INFINITESIMAL OPERATOR BASED METHODS FOR CONTINUOUS-TIME FINANCE MODELS

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Zhaogang Song

August 2011

© 2011 Zhaogang Song

ALL RIGHTS RESERVED

INFINITESIMAL OPERATOR BASED METHODS FOR CONTINUOUS-TIME FINANCE MODELS

Zhaogang Song, Ph.D.

Cornell University 2011

Continuous time Markov processes, including diffusion, jump-diffusion and Levy jump-diffusion models, have become an essential tool of modern finance over the past three decades. Nowadays, they are widely used in modeling dynamics of, for instance, interest rates, stock prices, exchange rates and option prices. However, data are always recorded at discrete points in time, e.g., monthly, weekly, and daily, although these models are formulated in continuous time. This feature makes most econometric inferential procedures developed for discrete time econometrics unsuitable for continuous time models and complicates the econometric analysis considerably. For example, estimators obtained by applying discrete time econometric methods to the discretized version of continuous time models are not consistent for a fixed sampling interval.

More seriously, although the maximum likelihood method is a very appealing econometric procedure due to its nice properties like efficiency, the transition density and hence likelihood function of most continuous time Markov models have no analytic expressions. This poses a serious impediment for the implementation of likelihood procedures. Many approaches have been proposed to deal with this problem but they either incur substantive computation burdens especially for multivariate cases or involve complicated approximation formulas with limited applicability. Consequently, there is a strong need for convenient econometric methodologies designed for continuous time mod-

els given discrete sampled data.

Unlike the transition density, the infinitesimal operator, as an important mathematical tool in probability theory, enjoys the nice property of being a closed-form expression of drift, diffusion and jump terms of the process. As a result, no approximated formulas or simulation based implementations are needed. Furthermore, it is equivalent to the transition density in characterizing the complete dynamics of the processes. Based on this convenient infinitesimal operator, this dissertation proposes a sequence of econometric procedures for continuous time Markov models with applications to affine jump diffusion (AJD) term structure models of interest rates. It is divided into four chapters.

In the first chapter, "Infinitesimal Operator Based Estimation for Continuous Time Markov Processes", I propose an estimation method based on the infinitesimal operator for general multivariate continuous-time Markov processes, which cover diffusion, jump-diffusion and Levy-driven jump models as special cases. A conditional moment restriction is first obtained via the infinitesimal operator based identification of the process. Then an empirical likelihood type estimator is constructed by a kernel smoothing approach. Unlike the transition density which is rarely available in closed-form, the infinitesimal operator has an analytic form for all continuous time Markov models. As a result, different from the maximum likelihood estimator (MLE) which involves either numerical or simulated transition densities, the proposed estimator can be conveniently implemented by plugging in parametric components of the models. Furthermore, I prove that the proposed estimator attains the semi-parametric efficiency bound for conditional moment restrictions models of Markov processes and hence is asymptotically efficient. Simulation studies show that the proposed estimator has good finite sample performances comparable to the MLE.

In the empirical application, I estimate Levy jump diffusion models for daily Euro/Dollar (2000-2010) and Yen/Dollar (1990-2000) rates. Results show that Levy jumps are important components in exchange rate dynamics and Poisson-type jump diffusion models cannot capture them.

In the second chapter, "Expectation Puzzles, Time Varying Conditional Volatility, and Jumps in Affine Term Structure Models", I study how jumps in interest rates, which are well documented in the literature, affect the term structure dynamics of the LIBOR-Swap curve in a multivariate AJD model. The motivation is that affine diffusion (AD) term structure models, as the major framework for interest rate dynamics, face two empirical challenges: first, they ignore well-documented jumps in interest rates as the state variables follow affine diffusions; second, they fail to capture simultaneously time variations in risk premiums implied by the violations of the "expectation hypothesis" and time variations in volatilities which are critical for pricing fixed-income derivatives. In this paper, I develop a multivariate AJD term structure model that overcomes these two challenges. Using LIBOR-Swap yields from 1990 to 2008, I estimate three-factor AJD models with infinitesimal operator methods and examine the contributions of jumps to term structure dynamics. I find that jumps are state dependent and negative. The risk premium is positive for jump size risk and negative for jump time risk, while the total jump risk premium is positive. Jump risk premiums lead to flexible time-varying market prices of risks without restricting time variations in conditional volatilities. As a result, two models in the three-factor AJD class capture time variations in both the risk premium and conditional volatility of LIBOR-Swap yields simultaneously.

In the third chapter (part of this chapter has been published as Song (2011) in Journal of Econometrics, 162-2, 189-212.), "A Martingale Approach for Testing

Diffusion Models Based on Infinitesimal Operator", I develop an omnibus specification test for diffusion models based on the infinitesimal operator instead of the transition density extensively used in literature. The infinitesimal operator based identification of the diffusion process is equivalent to a "martingale hypothesis" for the processes obtained by a transformation of the original diffusion model. My test procedure is then constructed by checking the "martingale hypothesis" via a multivariate generalized spectral derivative based approach which delivers an N(0,1) asymptotical null distribution for the test statistic. The infinitesimal operator of the diffusion process enjoys the nice property of being a closed-form function of drift and diffusion terms. Consequently, my test procedure covers both univariate and multivariate diffusion models in a unified framework and is particularly convenient for the multivariate case. Moreover, different transformed martingale processes contain separate information about the drift and diffusion specifications and about their interactions. This motivates me to propose a separate inferential test procedure to explore the sources of rejection when a parametric form is rejected. Simulation studies show that the proposed tests have reasonable size and excellent power performances. An empirical application of my test procedure using Eurodollar interest rates finds that most popular short-rate models are rejected and the drift mis-specification plays an important role in such rejections.

In the fourth chapter, "Estimating Semi-Parametric Diffusion Models with Unrestricted Volatility via Infinitesimal Operator", two generalized method of moments estimators are proposed for the drift parameters in both univariate and multivariate semi-parametric diffusion models with unrestricted volatility based on the infinitesimal operator. The first estimator is obtained by integrating out the diffusion function via the quadratic variation (co-variation), which is estimated by the realized volatility (covariance) in a first step using high frequency data. The second is constructed based on the separate identification condition and is actually applicable for a general instantaneous conditional mean model in continuous time, which covers the stochastic volatility and jump diffusion models as special cases. Simulation studies show that they possess fairly good finite sample performances.

BIOGRAPHICAL SKETCH

Zhaogang Song was born in Jinan, Shandong Province of China in June, 1979. He earned his B.S. degree in Management Science and Engineering from Shandong University in May, 2002. He then entered the graduate school of Shandong University and obtained his M.A. degree in Finance in May 2006. He continued his graduate studies of economics in the Department of Economics at Cornell University and will earn a Ph.D degree in Economics in May, 2011.

This document is dedicated to my wife, Qianqian Chen.

ACKNOWLEDGEMENTS

I am deeply indebted to my dissertation committee members Professor Yongmiao Hong, Professor Bob Jarrow, and Professor Nick Kiefer for many stimulating discussions.

I am particularly grateful to my committee co-chairs Professor Yongmiao Hong and Professor Bob Jarrow for guiding me through the fields of Econometrics and Finance respectively. They showed me what is a good research and taught me how to be a good researcher.

I really appreciate Professor Andrew Karolyi for spending his valuable time in reading my very preliminary papers and wrote numerous comments on every page. He even corrected many typos of mine. And He offered me so many help with my job market issues.

I wish to express my most sincere gratitude to Professor Haitao Li for his tireless encouragement and kindness. He is not only an outstanding mentor but also a good friend. He's talked to me on the phone for many many hours about research and my job search before he knew what I look like in Denver.

Last but not least, I want to thank my wife and parents for their love and support. My wife makes my life simple and delightful. She is my source of confidence during my job market period.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 Infinitesimal Operator-Based Estimation for Continuous-Time Markov Processes

Continuous-time Markov processes, including diffusion, jump-diffusion and Levy jump-diffusion models[1], have become an essential tool of modern finance over the past three decades. Such a modelling framework is able to simplify the financial optimization problem due to the elegant mathematics of stochastic calculus. Hence it has been widely used in analyzing the dynamics of, for instance, interest rates, stock prices, exchange rates and option prices. See Ait-Sahalia (1996a), Cox, Ingersoll and Ross (1985), Dai and Singleton (2003), Sundaresan (2000) and Wu (2008) for examples.

Although these models are formulated in continuous time, data are always recorded only at discrete points in time, e.g., daily, weekly, or monthly. This feature makes most econometric procedures developed in discrete-time econometrics unsuitable for continuous-time models and complicates the econometric analysis considerably. For example, Lo (1988) and Ait-Sahalia (2002) show that estimation methods in empirical finance which rely on the discretized version of the continuous time models, e.g., Chan, Karolyi, Longstaff and Sanders (1992) and Chapman and Pearson (2000), may result in inconsistent estimation when the sampling interval is considered as fixed. Consequently, there is strong need for consistent estimators of continuous-time models given discrete sam-

[1]The "jump-diffusion" without the qualifying "Levy" in this paper refers to jump-diffusion models driven by the compound Poisson process.

pled data, as can be seen from Sundaresan (2000): "The challenge to the econometricians is to present a framework for estimating such multivariate diffusion processes, which are becoming more and more common in financial economics in recent times. ... Recent developments in econometric theory give us considerable hope that more realistic multifactor continuous-time models can be estimated so that their practical implementation will be feasible. The development of estimation procedures for multivariate AJD processes is certainly a very important step toward realizing this hope". Note that the model considered in this paper includes both affine jump-diffusion (AJD in the quote) and non-affine processes and is hence more general than what Sundaresan (2000) suggests.

Due to its well-documented statistical properties such as efficiency, the MLE is very appealing and is preferred for estimating these continuous-time Markov models. However, it is difficult to implement MLE since the transition density and hence likelihood function have no analytic expressions in most cases[2]. Several approaches have been proposed to deal with this problem, including those in Lo (1988) by numerically solving the Fokker-Planck-Kolmogorov partial differential eqnarray, in Brandt and Santa-Clara (2002) by simulations based on Euler discretization, and in Elerian et al. (2001), Eraker (2001) and Johannes and Polson (2009) by Bayesian methods and MCMC. However, these procedures incur substantial computational burdens, especially for multivariate cases.

The seminal works of Ait-Sahalia (2002, 2008) and the follow up papers by Egorov, Li, and Xu (2003), Bakshi, Ju and Ou-Yang (2006), Schaumburg (2001), and Yu (2007) make the MLE practically feasible by providing closed-form approximation formulas based on Hermite polynomials. However, the deriva-

[2]See Wong(1964) for a list of rare diffusion models which do have the closed-form transition density.

tion of the recursively defined coefficients in the approximations involves "a multidimensional integral dependence and is seldom tractable outside of the constant elasticity of variance diffusion class" (Bakshi, Ju and Ou-Yang, 2006)[3], which may restrict the applicability. Moreover, the expressions for these coefficients are not very transparent and must be carefully re-derived for each specific model[4] even in univariate cases, not to mention the multivariate models for which a further Taylor expansion is needed to deal with the irreducibility (Ait-Sahalia, 2008).

Avoiding the cumbersome transition density, this paper proposes a convenient estimation method for general multivariate continuous-time Markov processes based on the infinitesimal operator. A conditional moment restriction is first obtained via the infinitesimal operator-based identification of the process. Then an empirical likelihood type estimator is constructed by a kernel-smoothing approach. The main advantage of the infinitesimal operator is that it enjoys a closed-form expression in terms of drift, diffusion and jump functions of a general Markov process. As a result, unlike the MLE for which the criterion function to be maximized must be constructed via numerical or simulated procedures, the proposed estimator can be conveniently implemented by plugging in the parametric terms of the models directly. Moreover, all popular continuous-time finance models, including diffusion, jump-diffusion and Levy jump-diffusion models, are unified in the same estimation framework by the infinitesimal operator which is defined for general continuous-time Markov pro-

[3]Bakshi, Ju and Ou-Yang (2006) propose a method refining the approximation formulas to depend directly on the original process and evaluating one- instead of multi-dimensional integrals. However, for multivariate diffusion models with this problem substantially harder, they "have been unable to work through the multivariate counterparts" (Bakshi, Ju and Ou-Yang, 2006) and hence their method are not guaranteed to work.

[4]For example, Ait-Sahalia and Kimmel (2010) have to develop the closed-form approximation formulas for each of the nine three-factor affine term structure models in Dai and Singleton (2000) case by case.

cesses. Hence in contrast to the approximated MLE of Ait-Sahalia (2002, 2008), the criterion function has the same form for all these models and does not need to be constructed case-by-case.

The estimation approach employed for the conditional moment restriction, obtained by the infinitesimal operator-based identification of the process dynamics, is adapted from the local empirical likelihood (LEL) method first proposed in Kitamura, Tripathi and Ahn (2004) (**KTA** hereafter). **KTA** develop this estimator in an I.I.D. setting and prove that it achieves the semi-parametric efficiency bound in Chamberlain (1987). As the second contribution of the paper, I extend the LEL method to a time series setup, i.e., m-th order Markov process, and show that it attains the semi-parametric efficiency bound provided most recently in Carrasco and Florens (2008) for dynamic models. Such an extension is of independent interest as a new estimation method for time series conditional moment restrictions models. It incorporates the information implied by the conditional moment restrictions efficiently and delivers an asymptotically efficient estimator without estimating the optimal instruments. Combined with the fact that the infinitesimal operator is equivalent to the transition density in characterizing the process dynamics, such an efficient use of the conditional moment restrictions implies that the proposed LEL estimator is close to the MLE in terms of the asymptotic efficiency.

Of course, by switching from the transition density to the convenient infinitesimal operator, we have to pay the price of approximating a numerical integral by a discrete sum due to the discrete nature of the data. However, this is very different from estimation methods based on the discretized version of the continuous-time models in that the it approximates a numerical integral in the estimator, while the latter discretize the stochastic model. To check the impact of the numerical errors, I conduct simulation studies of the proposed LEL estimator, with comparisons to the estimators from the discretized versions of the models using Euler discretization schemes. Results show that the proposed LEL estimator outperforms the Euler approximation schemes in situations relevant for financial models. Simulation studies are also conducted for comparisons with the MLE. Monte Carlo evidence reveals that the finite sample performances of the proposed LEL estimator are comparable to the MLE.

As a natural alternative to likelihood methods, estimation based on moment conditions is usually very simple and recurs as a major theme in econometrics, ranging from the classical OLS estimators to the extensively investigated generalized method of moments (GMM) (Hansen, 1982). However, obtaining exact moment conditions from conditional distributions of continuous-time models is not feasible since a closed-form expected value of certain functions cannot be obtained under the non-analytic transition density (see Ait-Sahalia (2007, Section 4.2.1) for detailed discussions). Different method-of-moment estimators have been proposed in the literature, either by simulation-based approaches like Gouriérous et al (1993), Gallant and Tauchen (1996) and Duffie and Singleton (1993), or by nonparametric smoothing-based minimum distance methods like Ait-Sahalia (1996a) and Bandi and Phillips (2007)[5]. But the former is computationally intensive and the latter is difficult to implement for multivariate models due to the "curse of dimensionality". The characteristic function (CF)-based

[5]Ait-Sahalia (1996a) mainly minimizes the distance between the parametric marginal density(it always has a closed-form) and its nonparamtric smoothed counterpart, while Bandi and Phillips (2007) minimize the distance between parametric drift and diffusion functions and their local nonparametric counterparts. The former is applicable to both high and frequency data and the latter only to high frequency data. At the same time, stationarity is required in the former and nonstationarity up to recurrence is allowed in the latter. In addition, separate estimation of drift and diffusion parameters, i.e., estimation of semi-parametric models, can be performed in the latter.

methods, including Carrasco et al. (2007), Chacko and Viceira (2003), Jiang and Knight (2002) and Singleton (2001) do derive exact moment conditions. But they are only applicable to cases where the CF has a closed form, which is true only for the class of affine models (Duffie, Pan and Singleton 2000; Chernov et al. 2003). My infinitesimal operator-based conditional moment restriction, by contrast, is valid and convenient for both affine and non-affine multivariate Markov processes. No simulations or numerical procedures are needed due to the analytic form of infinitesimal operator.

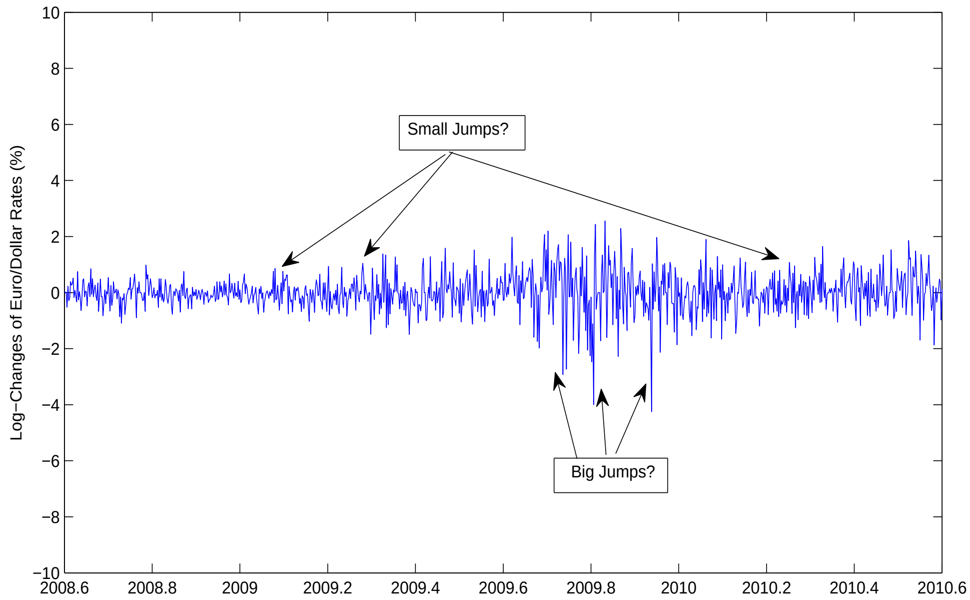
Earlier studies have also considered the infinitesimal operator in deriving exact moment conditions. For example, Kessler and Sørensen (1999) propose to use eigenfunctions of the infinitesimal operator to obtain moment conditions. However, the eigenfunctions cannot be computed in closed form except in special cases. Hansen and Scheinkman (1995) (**HS** thereafter) is, to the best of my knowledge, the only study which obtains exact moment conditions for general continuous-time Markov processes. It is the study most related to the present paper. **HS**'s derived moment conditions are in the form of expectations of the infinitesimal operator evaluated at a test function. A major drawback of the **HS** moment condition, as pointed out by Ait-Sahalia and Mykland (2008), is that for many models they cannot identify the parameters uniquely; in fact, only identification up to scale is ensured. The reason is that **HS** moment conditions are derived by the stationarity assumption of the process which does not uniquely characterize the process dynamics. In contrast, my infinitesimal operator-based conditional moment restriction, derived through the "martingale problem" as a complete identification condition of the models, is able to identify all the parameters uniquely. In addition, **HS** moment conditions are only developed for univariate models and time reversibility is required, while the proposed conditional moment restriction here is valid for general multivariate Markov processes whether time reversible or not.

As an empirical application, I apply the proposed LEL method to study Levy jump-diffusion models for exchange rate dynamics, which are important for understanding such financial issues as international trade and capital flows, international portfolio management, currency options pricing and foreign exchange risk management. Motivated by the fact that jumps can be generated by discontinuities in the arrival of "news" or by changes in monetary policies, jump-diffusion models driven by compound Poisson processes have been employed to model the exchange rate dynamics in recent years (Akgiray and Booth, 1988; Jorion, 1988; Bates, 1996). Figure 1.1 plots daily log changes in the Euro/Dollar rate from June 2008 to June 2010. Of special interest are the relatively infrequent but fairly large spikes, which are interpreted as jumps and can be modelled by jump-diffusions with compound Poisson processes.

However, Figure 1.1 also exhibits a large number of changes, which, although much smaller than those identified as Poisson-type jumps, are still relatively sizable. The natural next question is: are these changes truly jumps or time discrete variations from a Brownian motion? If they are just diffusive variations featured by Brownian motion, we can safely stay in the framework of Poisson-type jump-diffusion models in modeling the exchange rate dynamics. But what if these changes are indeed jumps with small magnitudes? In this case, compound Poisson process, as a finite-activity jump process which generates only a finite number of jumps within a finite time interval, is not able to capture such frequent and small jumps (see Li, Wells and Yu (2008) for details). As a result, we need the Levy process with infinite activity, whose jump compo-

Figure 1.1: **Time Series of Daily Log Changes of Euro/Dollar Rates**

This figure plots the daily Log changes in the Euro/Dollar rates from June, 2008 to June, 2010.



nent can accommodate an infinite number of small jumps within any finite time interval, to capture these frequent changes of small magnitude if they are truly Levy jumps[6].

A number of Levy jump models have been proposed in recent years (Carr and Wu, 2004a; Barndorff-Nielsen, 1988; Madan, Carr and Chang, 1998; Carr, Geman, Madan and Yor, 2002; Carr and Wu, 2003) and many studies have examined empirical performance of these models in capturing either the stock return or option prices dynamics (Chernov, Gallant, Ghysels and Tauchen, 2003;

[6]In this paper, Levy jumps represent the infinite-activity jump processes which exhibit a infinite number of jumps within any finite time interval.

Huang and Wu, 2003; Carr and Wu, 2003; Li, Wells and Yu, 2008, 2009). However, little has been done toward studying Levy jumps in exchange rate dynamics. Applying the proposed LEL method, I estimate Levy jump-diffusion models using daily Dollar/Yen exchange rates from 1988 to 2006. The models have CKLS-type specifications for the drift and diffusion functions and the variance-gamma (VG) and finite-moment log-stable (LS) Levy processes in Madan, Carr and Chang (1998) and Carr and Wu (2003) respectively for jump components. For comparison, I also consider compound Poisson processes for the jump specification. The estimation results show that all the jump parameters are significantly different from zero, implying the importance of accounting for the Levy jump behavior in the exchange rate dynamics. Similar to Johannes (2004) and Li, Wells and Yu (2008, 2009), I also examine the filtered jump variables by estimating the filtering distribution. The results show that the filtered Levy jump variables identify much more Levy-type small jumps than big Poisson-type jumps. These small jumps happen so frequently that they are most likely induced by normal market information flows like those related to transactions rather than by big economic announcements.

1.2 Expectation Puzzles, Time-Varying Conditional Volatility and Jumps in Affine Term Structure Models

Affine diffusion (AD) term structure models, in which the yields of zero coupon bonds are linear functions of the model state variables, are very popular among both practitioners and academics due to their convenient numerical and econometric tractability (Vasicek, 1977; Cox, Ingersoll and Ross, 1985, Chan, Karolyi, Longstaff and Sanders, 1992; Longstaff and Schwartz, 1992). Nowadays, when conducting empirical term structure studies, it is standard practice to employ the AD framework introduced in Dai and Singleton (2000), who characterize maximally flexible and empirically identifiable affine term structure models driven by pure diffusions; see Duffee (2002), Cheridito et al. (2007), Thompson (2008), and Ait-Sahalia and Kimmel (2010).

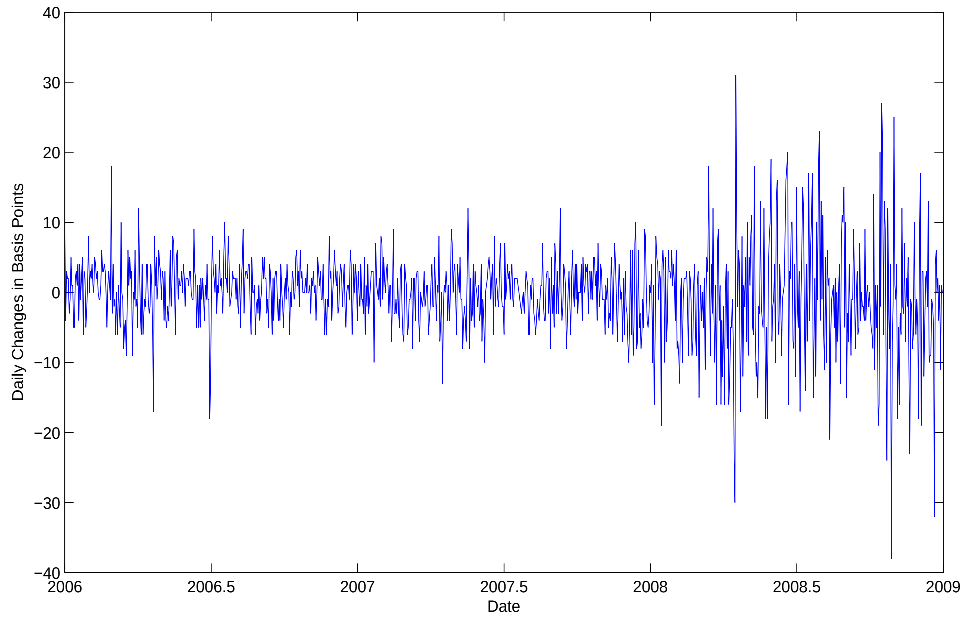
Yet, two empirical challenges still exist for AD term structure models. First, the AD framework assumes that interest rate movements are continuous, i.e., they follow pure diffusion processes. This approach may appear restricted in light of many recent studies (e.g., Johannes, 2004; Piazzesi, 2005; Jiang and Yan, 2009) which document the important role of jumps as "surprise elements" or unexpected discontinuous changes of large magnitude in interest rates. Figure 1.3 plots the daily changes in 2-year LIBOR-Swap rates from January 2006 to December 2008. Of special attention are the relatively infrequent but fairly large spikes, which are interpreted as jumps,[7] especially during the time period January 2008–December 2008. Johannes (2004) links similar jumps in the 3-month Treasury bill rate between 1991 and 1993 to the arrival of significant information regrading the current or future state of the economy. In the same spirit, the large spikes observed here are potentially connected to economic news, especially stories related to the recent financial crisis of 2008. Nonetheless, AD models cannot capture these jumps in interest rates, which contain important information about the market.

The second challenge facing AD term structure models, as documented in

[7] I conducted formal tests for the existence of jumps in LIBOR-Swap rates using the method of Chen and Song (2010). The results show that jumps exist in daily LIBOR-Swap yields of 3-month, 6-month, 9-month, 2-year, 3-year, 4-year, 5-year, 7-year, and 10-year maturities. See Appendix B for details.

Figure 1.2: **Time Series of Daily Changes in 2-year Swap Yields**

This figure plots daily changes in basis points of 2-year LIBOR-Swap rates from January 2006 to December 2008.



Dai and Singleton (2002), Duffee (2002), and Duarte (2004), is that AD models fail to capture time variations in risk premiums and conditional interest rate volatilities simultaneously. The empirical evidence that risk premiums exhibit time variations dates back at least to the compelling studies of Fama and Bliss (1987) and Campbell and Shiller (1991). In particular, Campbell and Shiller (1991) and Backus et al. (2001), among others, document an empirical pattern of violations of the "expectation hypothesis" for U.S. Treasury yields: Instead of resulting in the unity implied by the "expectation hypothesis," the regression coefficients of yield changes on yield spreads are negative, and increasingly so with longer maturities. This empirical pattern of deviations from the "expectation hypothesis," which centers on the assumption of constant risk premiums, captures in essence time variations in the interest rate risk premium (Dai and Singleton, 2003). Moreover, there exists substantial evidence that bond yields exhibit time-varying conditional second moments as well; see, for example, Ait-Sahalia (1996), Gallant and Tauchen (1998), and Andersen and Lund (1997). In fact, the term structure of unconditional volatilities of (changes in) yields tends to be hump-shaped for both U.S. Treasury securities and LIBOR-Swap curves (Litterman, Scheinkman, and Weiss, 1991; Dai and Singleton, 2000, 2003).

These documented time variations in the risk premium and conditional volatility are key components of pricing fixed-income securities: The former capture the bond risk premia critical for pricing bonds of differing maturities; the latter are particularly important for the reliable valuation of many fixed-income derivatives such as interest rates swaptions, caps and floors. Therefore, risk premium and conditional volatility time variations have been treated as two stylized facts and descriptive statistics that an empirically successful dynamic term structure model should match (Dai and Singleton, 2003). However, there is strong evidence that AD models do not match these two stylized facts simultaneously. In particular, Dai and Singleton (2002) and Duffee (2002) find that AD models that are flexible enough to capture time variations in the risk premium are wholly incapable of generating any time variation in interest rate volatilities. Therefore, serious tension exists in AD models between matching the first- and second- order moments of the interest rate data.[8]

In this study, I develop a multivariate AJD term structure model, which augments AD models in Dai and Singleton (2000) by allowing jumps in the model

[8]This mean-volatility tension is shown to exist in other models outside of the affine class as well; see Duarte (2004) and Ahn, Dittmar and Gallant (2002).

risk factors, to meet the two empirical challenges. Following Pan (2002) and Jarrow, Li, and Zhao (2006), we first specify the risk factor dynamics under both the physical and risk-neutral measures. Then jump risk premiums are calculated as the difference of the physical and risk-neutral dynamics of jumps. Adding jumps into the model state variables is directly motivated by the first empirical challenge with AD models, namely the inability to capture observed jump behaviors on the part of interest rates. Using daily LIBOR-Swap rates from 1990-2008, I document important features of jumps in interest rates, including the signs of jumps and jump risk premiums. Furthermore, I find that the second empirical challenge, namely the tension between matching the time-variation in risk premiums and matching time-varying conditional volatility, is also met in the proposed AJD model.

To characterize time-varying risk premiums, I first run the regressions in Campbell and Shiller (1991) and Backus et al. (2001) of yield changes on yield spreads and document the empirical pattern of the violation of the ”expectation hypothesis” for daily LIBOR-Swap yields from 1990 to 2008. The primary motivation for choosing swap yields instead of Treasury yields, which are mostly used in previous studies of the ”expectation hypothesis” (Fama and Bliss, 1987; Campbell and Shiller, 1991; Backus et al., 2001; Dai and Singleton, 2002), is the availability of high-frequency daily data that is more relevant for studying jumps (see Section 2 for more advantages of LIBOR-swap yields). The estimated regression coefficients decrease with longer maturities, change from positive values for rates with maturities less than two years to negative values for rates with maturities larger than two years, and are very close to zero overall in magnitude, with values ranging between -0.0051 and 0.0096. This empirical pattern of deviations from the unity line, which is implied by the ”expectation hypothesis," characterizes time variations in the risk premium on LIBOR-Swap yields.

To capture time variations in interest rate volatilities, the term structure of volatilities is computed for LIBOR-Swap rates and a hump shape is found with the hump occurring at around the 9-month to 1-year maturity range.[9] Moreover, I follow Dai and Singleton (2003) to estimate a GARCH(1,1) model for the yields and a high degree of persistence is found for yields with all available maturities. These two documented stylized facts are strong evidence of time variation and persistence in yield volatilities (Dai and Singleton, 2003).

Then, following the term structure literature (Dai and Singleton, 2002, 2003; Duffee, 2002) and motivated in part by Litterman and Scheinkman (1991) who find that three principal components neatly capture over 90% of variations in U.S. treasury yields,[10] I employ three-factor models to match the documented patterns above of time variations in both risk premiums and conditional volatilities. Both three-factor AD and AJD models are estimated, adopting the "essentially" affine specification for market prices of diffusive risk (Duffee, 2002). Moreover, I specify the jump risk premium to compensate for both the jump size uncertainty and the jump time risk. The data used for estimation are daily LIBOR-Swap rates with 6-month, 2-year, 3-year, 5-year, 7-year and 10-year terms of maturity from August 13, 1990 to December 31, 2008. Similar to Dai and Singleton (2002), I calculate the relevant population and sample versions of regression coefficients and check whether they match the time variation in risk

[9]Dai and Singleton (2000), using weekly LIBOR-Swap yields from 1987 to 1996, find that the hump happens at around the 2-year to 3-year range of maturity.

[10]Dai and Singleton (2000) find that LIBOR-Swap yields and U.S. Treasury yields have similar distributional characteristics, including the principal components and yield volatilities although the institutional structures of the two markets are different. Hence the focus here on three-factor models for the LIBOR-Swap curve can also be justified by Litterman and Scheinkman (1991).

premiums. In addition, time series of LIBOR-Swap rates are simulated from the models evaluated at their estimated parameter values. Then sample variances are computed and compared with the empirical pattern to see whether the models can capture the time variation in interest volatility (Dai and Singleton, 2000; Piazzesi, 2005).

Results for the AD models here using LIBOR-Swap yields are very similar to those found by Dai and Singleton (2002) using U.S. Treasury yields. That is, the only model in the three-factor AD class that can capture the time variation in risk premiums, the Gaussian-model, exhibits no time-variation in conditional volatility specifications. In sharp contrast, results for the AJD models show that two models in the three-factor AJD class simultaneously match time variations in both the risk premium and conditional volatility. In fact, all four models in the three-factor AJD class, including the one with the most flexible time-varying conditional volatility specifications, can closely match the time-varying risk premium of the LIBOR-Swap curve. Therefore, the empirical evidence suggests that the tension between matching the first- and second- order moments of interest rates, which exists in the AD framework, is eliminated in AJD models.

By analyzing the structures of affine term structure models, I provide theoretical support for the empirical success of AJD models in simultaneously capturing time variations in both risk premiums and conditional volatilities. The key is the jump risk premium, which generalizes the market prices of risk without restricting the time-varying conditional volatility. Specifically, the market prices of risk in "essentially" AD models are tied to the conditional volatility due to the affine structure: As state factors in the conditional volatility specification generate more flexible time-varying conditional volatility, fewer elements in the market prices of risk will be able to switch signs over time, producing more restrictive time-varying risk premiums. In contrast, for AJD models, the jump risk premium enables every element in the market prices of risk to change signs over time while imposing no single restriction on conditional volatility. That is, the tight link between the market prices of risk and the time-varying conditional volatility in AD models is indeed broken up by the introduction of jump risk premiums, which account for the documented empirical success of the AJD models.

The estimation results in this paper also allow us to obtain detailed information about jumps in interest rates as a means of addressing the first empirical challenge. First, I find that most parameters in the jump specifications are significantly different from zero, implying the importance of modeling jumps in capturing the dynamics of the LIBOR-Swap curve. In particular, the statistical significance of jump intensity parameters shows that jump arrivals are state dependent,[11] implying a certain level of predictability with respect to current market conditions for the frequency of future large changes in yields. Second, overwhelming evidence of negative jumps in the state variable dynamics is found under both physical and risk-neutral measures. As suggested in Jarrow, Li and Zhao (2007), this may reflect investors' fears of a market crash such as that of 1987. Third, the risk premium is positive for jump size risk and negative for jump timing risk. But since most jumps are negative, the total jump risk premium is positive. The large discrepancy between jump intensities under the physical and risk-neutral measures is similar to that shown for jump sizes in Jarrow, Li and Zhao (2007) using LIBOR rates and in Pan (2002) using S&P 500

[11]In contrast, Piazzesi (2005), in a study of interactions between bond yields and policy decisions by the Federal Reserve, models the jumps as deterministic by linking jump intensities directly to the meeting calendar of the Federal Open Market Committee.

index options. By analogy, this might be explained by a huge jump risk premium.

So far, ”state variables with jumps have received relatively less attention in the empirical literature on DTSMs”[12] (Dai and Singleton, 2003) and ”there is little work of jump-diffusion term structure models” (Johannes and Polson, 2009), especially multifactor AJD term structure models. Many studies have incorporated jumps in modeling term structure dynamics, such as Ahn and Thompson (1988), Das (2002), Anderson et al. (2004), Jarrow, Li and Zhao (2007), and Zhou (2001). However, these studies focus only on special cases of the general AJD framework of this paper[13] in the sense of being maximally flexible and econometrically identifiable (Dai and Singleton, 2000). While some theoretical work, such as Chacko and Das (2002), develops a general approach for pricing interest rate derivatives in the AJD framework, many important questions have not been answered empirically: ”Do multiple factors jump, or is it only the short rate? Does the market price diffusive and jump risks differently in the term structure? How do predictable jumps affect the term structure?” (Johannes and Polson, 2009). The first two questions of Johannes and Polson (2009) are related to the first empirical challenge stated at the beginning of the paper and the last one to the second. This paper answers them all: Multiple risk factors do jump; the jump risk premium is positive though the diffusive (volatility) risk premium is negative; the predictable jumps are crucial for enabling AJD models to simultaneously capture time variations in both the risk premium and conditional volatility.

[12]DTSMs stands for ”dynamic term structure models” in Dai and Singleton (2003)

[13]Piazzesi (2005), Cheng and Scaillet (2007), and Jiang and Yan (2009) cannot be covered by this AJD framework directly because their models have a quadratic component in the state variable vector. However, since jumps can happen only in the dynamics of affine state variables, these models can also be regarded as special cases of the AJD framework as far as the jumps are concerned.

One possible reason that there have been so few studies of jumps in the term structure dynamics of interest rates may be the lack of convenient econometric methods. It is well known that estimating continuous-time finance models is very challenging since data are always recorded only at discrete points in time although the models are formulated in continuous time. This feature makes most econometric procedures developed for discrete-time econometrics unsuitable for continuous time-models and complicates the econometric analysis considerably.[14] More seriously, the likelihood function of most continuous-time Markov models have no analytic expressions, which represents a serious impediment to the implementation of the statistically appealing maximum likelihood estimation (MLE).[15] The estimation method in this study is adapted from Chapter 2, which is originally developed for multivariate continuous-time Markov models without unobservable state variables, to the affine term structure models with latent risk factors. It avoids the cumbersome transition density and depends on the infinitesimal operator, which features the nice property of being a closed-form expression of drift, diffusion and jump terms in a general Markov process. No approximated formulas or simulation-based implementations are needed and the method is numerically convenient. Furthermore, the infinitesimal operator is equivalent to the transition density in characterizing the complete dynamics of the processes. Consequently, the infinitesimal opera-

[14]For example, Lo (1988) and Ait-Sahalia (2002) show that estimation methods in empirical finance that rely on the discretized version of continuous-time models, e.g., Chan, Karolyi, Longstaff and Sanders (1992) and Chapman and Pearson (2000), may result in inconsistent estimation when the sampling interval is considered as fixed.

[15]Various methods for implementing MLE have been proposed, including Lo (1988) by solving numerically the Fokker-Planck-Kolmogorov partial differential eqnarray, Brandt and Santa-Clara (2002) by simulations based on Euler discretization, Elerian et al. (2001), Eraker (2001) and Johannes and Polson (2009) by Bayesian methods and MCMC, and Ait-Sahalia (2002, 2008) by a closed-form approximation through Hermite polynomials. The characteristic function-based methods, including Chacko and Viceira (2003), Jiang and Knight (2002) and Singleton (2001) can also be used to estimate continuous-time finance models with closed-form characteristic functions. Recently, Czellar, Karolyi and Ronchetti (2007) proposed a simulation-based indirect robust generalized method of moments estimation method.

tor method employs the same information set of process dynamics as that used by MLE and the empirical results in this paper are not specific to estimation methods like those matching only certain moments of the interest rates processes. Such an estimation method is of independent interest as can be seen from Sundaresan (2000): "Recent developments in econometric theory give us considerable hope that more realistic multifactor continuous-time models can be estimated so that their practical implementation will be feasible. The development of estimation procedures for multivariate AJD processes is certainly a very important step toward realizing this hope."

1.3 A Martingale Approach for Testing Diffusion Models Based on Infinitesimal Operator

Diffusion models have proven to be mostly successful in finance over the past three decades in modeling the dynamics of for instance interest rates, stock prices, exchange rates and option prices. Since economic theories usually do not suggest any concrete functional form for the processes, the choice of a model is somewhat arbitrary and a great number of parametric diffusion models have been proposed in the literature, see for example Ait-Sahalia (1996a), Ahn and Gao (1999), Chan, Karolyi, Longstaff and Sanders (1992), Cox, Ingersoll and Ross (1985), and Vasicek (1977). However, model misspecification may yield misleading conclusions about the dynamics of the process and result in large errors in pricing, hedging and risk management. The development of reliable specification tests for diffusion models is therefore very necessary to tackle such problems.

In this study, I develop an omnibus test for the specification of diffusion models based on the infinitesimal operator which is a complete characterization of the whole dynamics of the process alternative to transition density used by Ait-Sahalia, Fan and Peng (2009), Chen, Gao and Tang (2008), and Hong and Li (2005). Through the celebrated "martingale problem" developed by Strook and Varadhan (1969), the infinitesimal operator based martingale characterization of a diffusion process is obtained such that the identification of the diffusion process is equivalent to a "martingale hypothesis" for the processes transformed from the original diffusion process. I then check the "martingale hypothesis" via a multivariate generalized spectral derivative approach, which is an extension of Hong (1999) for univariate time series processes. Such a test is particularly powerful against alternatives with zero autocorrelation but a nonzero conditional mean and has a convenient one-sided $N(0,1)$ asymptotic distribution. The infinitesimal operator of the diffusion process enjoys the nice property of being a closed-form expression of drift and diffusion terms. This makes my test procedure feature many good properties which will be discussed in the following.

Ait-Sahalia (1996a) developed probably the first nonparametric test for univariate diffusion models by comparing the model-implied stationary density (or transition density) with a smoothed kernel density estimator based on discretely sampled data. Hong and Li (2005) observed that when a diffusion model is correctly specified, the probability integral transform (PIT) of data via the model-implied transition density is *i.i.d* $U[0,1]$. Then an omnibus test is proposed by checking the joint hypothesis of *i.i.d* $U[0,1]$ though a smoothed kernel estimator of the joint density of the probability integral transform series. As by-products of the Efficient Method of Moments (EMM) algorithm, a χ^2 test for model misspecification and a class of appealing diagnostic t-tests to gauge possible sources of model failure are proposed in Gallant and Tauchen (1996), which are applicable to general continuous time models. The idea is to match the model-implied moments to those implied by a semi-nonparametric (SNP) transition density for observed data.

Many other tests have appeared recently for univariate diffusion models based on the transition density directly. Both Ait-Sahalia, Fan and Peng (2009) and Chen, Gao and Tang (2008) proposed tests by comparing the model-implied parametric transition density and distribution function to their nonparametric counterparts, with latter using a nonparametric empirical likelihood approach. Corradi and Swanson (2005) introduced two bootstrap specification tests for diffusion processes. The first, for one-dimensional case, is a Kolomogorov type test based on comparison of the empirical cumulative distribution function(CDF) and the model-implied parametric CDF. The second, for multidimensional or multifactor models characterized by stochastic volatility, compares the empirical distribution of the actual data and that of the (model) simulated data. Noticing most of the tests for diffusions apply only for the univariate case, Chen and Hong (2010) considered a test for multivariate diffusion models based on the conditional characteristic function (CCF) which is the Fourier transform of the transition density.

Different from all the tests above, my proposed test procedure is based on the so-called infinitesimal operator which can completely characterize the dynamics of the underlying continuous time process. Intuitively speaking, the infinitesimal operator captures the limiting behavior of the conditional movement and hence the whole dynamics of the process since the time goes continuously.

Several properties of the infinitesimal operator and the relevant advantages of the proposed test are on the way. First, alternative to the transition density, the infinitesimal operator is also able to completely identify the dynamics of the diffusion process. Consequently, my infinitesimal operator based test can pick up effectively the misspecified models that have a correct stationary density which Ait-Sahalia's (1996a) marginal density-based test may easily pass over. It hence significantly improves the size and power performance of the marginal density-based test. In such a sense, my test is omnibus, unlike Gallant and Tauchen's (1996) EMM tests which, as Tauchen (1997) points out, are not consistent against all model misspecifications because they are based on a semi-nonparametric score function rather than the transition density itself.

Second, the infinitesimal operator has always a closed-form expression in terms of drift and diffusion functions. In contrast, it is well known that the transition density of most continuous time models has no closed form. As a result, some techniques to approximate the transition density are required in the transition based tests(see Hong and Li (2005) and Ait-Sahalia, Fan and Peng (2009)), for example, the simulation methods of Brandt and Santa-Clara (2002), the Hermite expansion approach of Ait-Sahalia (2002), or for affine diffusions, the closed-form approximation of Duffie, Pedersen, and Singleton (2003) and the empirical characteristic function approach of Singleton (2001) and Jiang and Knight (2002). Although the asymptotic distribution of some tests (like Hong and Li(2005)) is not affected by the estimation uncertainty, the use of the transition density may not be computationally convenient and may affect the finite-sample performance of the test. However, my infinitesimal operator based test requires nothing except the drift and diffusion terms. No approximation techniques are needed and the test is easy to implement and computationally convenient.

Third, the closed-form expression of the infinitesimal operator for multivariate cases are similar to and as simple as that for univariate cases. Hence, the proposed test is particularly convenient for checking multivariate diffusion models, which is fairly difficult by other methods. For example, Hong and Li's (2005) approach cannot be extended to a multivariate context directly because the PIT of data with respect to a model-implied multivariate transition density is no longer *i.i.d* $U[0,1]$, even if the model is correctly specified. Although they propose to evaluate multivariate models using the PITs for each state variable which is valid by partitioning the information set appropriately, it may fail to detect misspecification in the joint dynamics of state variables. In particular, their test may easily overlook misspecification in the conditional correlations between state variables, which are known to be important in term structure literature (Dai and Single, 2000). Chen and Hong (2010) do have the ability to check multivariate diffusion models but their test depends crucially on the availability of closed-form CCF. For the proposed test here, univariate and multivariate diffusions are unified in the same framework and no additional steps are necessary for multivariate cases.

Fourth, the infinitesimal operator based martingale characterization of diffusion models can reveal separate information about the specification of drift and diffusion terms or even their interactions. This is a property which no other approaches enjoy so far. Although other methods are also available to check the specification of the drift or diffusion terms by nonparametrically smoothing only one of them, the infinitesimal operator based martingale characterization proposed in this study brings up this type of information in an essential way.

This motivates me to suggest a separate inference test to determine the sources when rejection of a parametric form happens. Not only is my test convenient for multivariate models which are difficult for other methods like Li (2007) and Kristensen (2008), but it is constructed in exactly the same framework as the proposed test for joint dynamics. In other words, a unified procedure is developed to first check the specification for the joint dynamics and then gauge sources of rejection in order to build a more accurate model for financial variables.

This paper is also related to the literature of operator methods for continuous time processes (see Ait-Sahalia, Hansen and Scheinkman(2004) for a survey), including the GMM-type and estimating eqnarray-type estimators in Hansen and Scheinkman (1995) and Kessler and Sorenson (1996) respectively, identification problem in Hansen, Scheinkman and Touzi (1998), semi-group pricing theory in Hansen and Scheinkman (2003), and the test in Kanaya (2007). Different from the econometric studies above using operator methods, the infinitesimal operator is utilized here via the martingale characterization which can also be extended to construct estimators of diffusion models and tests of whether a continuous time process is a diffusion generically like Kanaya (2007). Several nice advantages over the existing studies are expected to come up with the properties of the infinitesimal operator based martingale characterization including, for example, the complete identification of the diffusion process unlike Hansen and Scheinkman's(1995) identification up to scale, the closed-form expressions unlike the eigenfunctions used in Kessler and Sorenson (1996) and Hansen, Scheinkman and Touzi (1998), and convenient applications to multivariate diffusions unlike Kanaya (2007) which is only for univariate cases. These research are being investigated and will be reported soon.

1.4 Estimating Semi-Parametric Diffusion Models with Unrestricted Volatility via Infinitesimal Operator Based Characterization

Continuous time models have proven to be mostly successful in finance over the past three decades for modeling the dynamics of, for instance, interest rates, stock prices, exchange rates and option prices. Among the existing studies, diffusion models may be the most extensively employed and studied. The elegant mathematical tool for solving many important problems in finance, i..e., stochastic calculus, serves as an important reason for their popularity. While economic theories have implications about the relationship between economic variables, they usually do not suggest any concrete functional form for the processes; the choice of a model is somewhat arbitrary. Consequently, many parametric diffusion models have been proposed in the literature(Ait-Sahalia 1996a; Ahn and Gao 1999; Chan, Karolyi, Longstaff and Sanders 1992; Cox, Ingersoll and Ross 1985; and Vasicek 1977). However, this fully parametric approach is subject to the risk of mis-specification which could lead to misleading conclusions about pricing, hedging and risk management.

A full nonparametric diffusion model is evidently robust to mis-specifications and has received much attention recently; see Bandi and Phillips(2003), Stanton (1997), Chapman and Pearson (2000), and so on. However, a price has to be paid for such a robustness, usually in terms of the precision and convenience: nonparametric estimation has a low convergence rate and is difficult to be applied for multivariate models due to the notorious "curse of dimensionality".

A semi-parametric approach perfectly fits the framework of diffusion models since they are fully characterized by the so-called drift and diffusion functions, which capture instantaneous changes in conditional mean and volatility of the underlying process respectively. It provides a good compromise between the robustness of model specifications and convenience of practical implementation and has been employed in finance literature for modeling and pricing issues. For example, Ait-Sahalia(1996b) utilizes a semi-parametric diffusion model with a linear parametric drift and nonparametric diffusion to price the interest rate derivatives. For modeling the short-term interest rate, Conley et al.(1997)'s model can be roughly regarded as semi-parametric while Kristensen(2004) formally takes a semi-parametric approach. In this study, we shall consider semi-parametric diffusion models with parametric drift and nonparametric diffusion components, focusing on the consistent estimation of the drift parameters. Specifically, two GMM type estimators of drift parameters will be proposed for both univariate and multivariate cases. The conditional moment restriction, through which the estimators are constructed, follows from an infinitesimal operator based characterization of diffusion processes, which is first proposed in Song (2011) in constructing a specification test for parametric diffusion models.

There are many nice properties enjoyed by the infinitesimal operator, among which two features may be the most exciting: first, it is equivalent to transition density in terms of capturing the complete dynamics of the diffusion process; second, it has a closed-form expression in terms of drift and diffusion terms in an essentially separate manner. The former enables us lose no information about the dynamic probability law of the process and the latter is the cornerstone based on which we estimate the drift parameters robust to diffusion mis-specification. Two estimators are proposed via the infinitesimal operator based characterization in this paper. The first estimator is obtained by integrating out the diffusion function via the quadratic variation(covariation), which is estimated by the realized volatility(covariance) in a first step using high frequency data. The second estimator is constructed based on the separate identification condition and is actually applicable for a general instantaneous conditional mean model in continuous time proposed by Park(2008), which covers the stochastic volatility and jump diffusion models as special cases. Our estimators for both univariate and multivariate models are unified in the same framework and particularly easy-to-implement. We conduct a comprehensive simulation study and find that both of the two proposed estimators possess fairly good finite sample performances.

Although the semi-parametric approach to diffusion models are important, especially when the researchers have some prior beliefs about the shape of either drift or diffusion functions but not both, only a few econometric inference procedures exist in the literature and a large fraction of them is only focused on the consistent specification testing of either drift or diffusion functions(Li, 2007; Corradi and White, 1999; Kristensen, 2008b; Fan and Zhang, 2003). Recently, there have been several estimators proposed to estimate semi-parametric diffusion models consistently[16]. For example, Ait-Sahalia(1996b) considers a semi-parametric diffusion model with linear drift and nonparametric diffusion function. Due to the linearity of the drift, the parameters can be estimated

[16]Most recently, Phillips and Yu (2009) propose a two-stage approach for estimating drift and diffusion parameters separately. In the first stage, the diffusion parameters are estimated based on the equality of integrated diffusion term and quadratic variation. In the second stage, an in-fill(the sampling frequency converges to zero) likelihood function is maximized to obtain estimators for drift parameters. However, their estimator is only applicable for semi-parametric diffusion models with unrestricted drift instead of for those with unrestricted volatility since the in-fill likelihood function requires a known form of the diffusion function. Hence, it is not applicable for the cases we consider in this paper.

consistently by an OLS procedure. However, this OLS estimator is not valid under a general nonlinear drift specification(see Kristensen(2008a, p.6)). Kristensen(2008a) studies a semi-parametric diffusion model with a general drift specification and proposes to estimate the parameters using Pseudo-MLE via the link among stationary density, drift and diffusion functions. Since the likelihood function is not available in closed-form, his estimator has to be implemented by simulation or approximation methods and is thus computationally demanding. The two estimators proposed here, also for the semi-parametric diffusion model with a general parametric form for drift, however, are depending on the closed-form infinitesimal operator, thus excluding the need for simulation or approximation methods.

Furthermore, the link among stationary density, drift and diffusion function, which both Ait-Sahalia(1996a) and Kristensen(2008a) rely on, does not hold in multivariate cases. Therefore, their estimators do not apply for general multivariate diffusion models. In contrast, both of the two estimators we propose do work for multivariate cases and are hence more applicable. Actually, they may be the first consistent estimators for general multivariate semi-parametric diffusion models according to our best knowledge. Another issue is that the nonparametric estimation methods employed in Ait-Sahalia(1996a) and Stanton(1997) have been seriously challenged by Chapman and Pearson(2000) who point out that these nonparametric methods are subject to the finite sample bias due to the truncation of a distribution and are very unreliable. Since our approach is completely parametric, the proposed estimator involves no user-chosen number, enjoys the $\sqrt{n}$-convergence rate, applicable to multivariate case in a simple way, and is free of the finite-sample bias discussed above for nonparametric methods. Henceforth, the finite sample performance is expected to be better than those based on nonparametric estimation. Of course, our first estimator using realized volatility(covariance) requires the sampling interval to shrink and is only applicable to high-frequency data while both Ait-Sahalia(1996a) and Kristensen(2008a) applicable to low frequency data with a fixed sampling frequency. But the second estimator we propose based on the separate identification condition does work for data with a fixed sampling interval.

Bandi and Phillips(2007) propose to estimate semi-parametric diffusion models by minimizing the distance between parametric drift and diffusion functions and their local nonparametric counterparts, which is only applicable to high frequency data. A nice feature of their methods is that nonstationarity up to recurrence is allowed. However, their estimator requires an additional nonparametric estimation which usually involves a user-chosen bandwidth number. The sensitivity of the estimator to the choice of the bandwidth has to be evaluated and hence the estimator has to be used in caution. Moreover, it is well known that the nonparametric estimation has a lower convergence rate than parametric estimation. This makes the convergence of the estimator to the true parameter slowly and the finite sample performance may not be satisfying In addition, due to the "curse of dimensionality", this approach is difficult to be extended for multivariate diffusion models.

Recently, Park(2008) proposes a so-called "conditional mean model of instantaneous change for a given stochastic process"[17] and the identification of the model is equivalent to a martingale property in continuous time, which co-

[17]Caution is needed for these terminologies. The instantaneous conditional mean for continuous time stochastic processes are different from the conditional mean for discrete time models. As discussed earlier, for instance, in a general diffusion process, the conditional mean of $X_{t+\Delta}$(Δ is the sampling frequency) given X_t is usually a function not of drift solely but of both drift and diffusion terms jointly. See Ait-Sahalia(1996a) for more discussions.

incides with the separate identification condition we employ to construct our second estimator. This instantaneous conditional mean model is more general than Markovian diffusion models and actually covers as special cases both non-Markovian stochastic volatility models and processes with jumps like jump-diffusion models. However, the estimator proposed in Park(2008) is based on a time change technique transforming calender time to volatility time (quadratic variation clock). Since the time change requires estimating the quadratic variation, Park's (2008) estimator needs sampling interval shrink to zero and requires high frequency data. In addition, continuity of the sample path is also required for time-change which rules out jumps and limits the applicability of the model. In contrast, our second estimator via separate identification only depends on the conditional moment restriction implied by the martingale property and is applicable to both high and low frequency data and especially models with jumps such as the jump diffusion model. Moreover, Park(2008) only considers a univariate model since the time change technique does not apply for multivariate cases while our second estimator is able to estimate the multivariate version of the instantaneous conditional mean model, enlarging its applicability greatly.

CHAPTER 2

INFINITESIMAL OPERATOR-BASED ESTIMATION FOR CONTINUOUS TIME MARKOV PROCESSES

2.1 Infinitesimal Operator-Based Conditional Moment Restrictions

2.1.1 Infinitesimal Operator

The model we consider is a multivariate Markovian semimartingale defined by the following stochastic differential eqnarray (SDE) on the filtered probability space $(\Omega, \mathfrak{F}, \{\mathfrak{F}_t\}_{t\geq 0}, P)$:

$$dX_t = b(X_t; \theta)dt + \sigma(X_t; \theta)dW_t + dJ_t \tag{2.1}$$

where W_t is a $d \times 1$ standard Brownian motion in $\mathbb{R}^d$, $b : E \subset \mathbb{R}^d \to \mathbb{R}^d$ is a drift function (i.e., instantaneous conditional mean), $\sigma : E \to \mathbb{R}^{d\times d}$ is a diffusion function (i.e., the instantaneous conditional standard deviation), and $\Theta \subset \mathbb{R}^p$ is a finite-dimensional parameter space. The jump process J_t can be of a Poisson-type with jump arrival intensity $\lambda(X_t, \theta)$ and random jump size vector ξ_t, which is independent of $\mathfrak{F}_{t-}$ and has probability density $\nu(\cdot, \theta) : \mathbb{R}^d \to \mathbb{R}$. It can also be a pure jump Levy process with infinite activity which accommodates an infinite number of jumps within any finite time interval and is characterized by the triplet $\left(\mu, \sigma_L^2, \pi(dx, \theta)\right)$, usually referred to as Levy characteristics of an infinitely divisible distribution. The Levy measure $\pi(dx; \theta)$ captures the jump structure of the process (see Section 2.4.1 for a more detailed introduction of Levy processes) and satisfies

$$\int_{\mathbb{R}_0^d} \pi(dx;\theta) = \infty, \mathbb{R}_0^d = \mathbb{R}^d \backslash \{0\}$$

Therefore, the model framework in (2.1) is general enough to cover most popular continuous-time financial models (diffusion, jump-diffusion, and Levy-type jump models) for option pricing, term structure of interest rates, and exchange rate dynamics. See Sundaresan (2000) for a general survey of continuous-time finance, Dai and Singleton (2003) for term structure models and Wu (2008) for Levy-type models.

For the Model (2.1), conditions are needed to ensure that the parametric dynamics are well-defined. Following the literature (Ait-Sahalia, Fan and Peng, 2009; Ait-Sahalia and Mykland, 2003, 2008; Yu, 2007), we impose directly:

Assumption 2.1.1: The specification of $b(\cdot)$, $\sigma(\cdot)$, and J_t is such that the model (2.1) admits a unique solution and satisfies the smoothness and boundary behavior necessary to prevent the process from exploding.

Primitive conditions in terms of $b(\cdot)$, $\sigma(\cdot)$, and J_t that ensure Assumption 2.1.1 can be found in Ait-Sahalia (1996a, b; 2002; 2008), Ait-Sahalia and Myklan (2004), Brandt and Santa-Clara (2002), and Yu (2007). See Karatzas and Shreve (1991) and Rogers and Williams (2000) for proofs. The set E in (2.1) is often called state space, and we let $\mathcal{B}(E)$ be the Borel $\sigma-$field such that $(E, \mathcal{B}(E))$ is a measurable space. Under usual regularity conditions, $\{X_t\}$ is a continuous-time Markov process with transition function $P(t, x, \Gamma) \equiv P(X_t \in \Gamma | X_0 = x)$, which is equal to the probability that X_t, starting from the point x at the beginning time, is in the set Γ at time t. The Markov property is characterized by the so-called Chapman-Kolmogorov eqnarray: for $s, t \geq 0$, $x \in E$ and $\Gamma \in \mathcal{B}(E)$,

$P_{t+s}(x,\Gamma) = \int_E P_s(x,dy)P_t(y,\Gamma)$. An alternative and equivalent characterization is in terms of the induced family $\{P_t\}$, which is a set of positive bounded operators defined by:

$$P_t f(x) \equiv (P_t f)(x) = \int_E P_t(x,dy)f(y)$$

with norm less than or equal to 1 on $b(\mathcal{B}(E))$ (bounded and $\mathcal{B}(E)$-measurable functions). The Markov property can then be expressed as the so-called semi-group property, i.e., $P_s P_t = P_{s+t}$, for any $s,t \geq 0$.

The interaction between the semi-group property and sample-path property of a process can be used to define a special class of processes called Feller processes, including (2.1) as a special case. Let $C_0 = C_0(E)$ be defined as the space of real-valued, continuous functions on E which vanish at infinity, i.e., $\lim_{|x|\to\infty} f(x) = 0$, equipped with the sup-norm $\|f\| \equiv \sup_{x\in E} f(x)$. By Rogers and Williams (2000, Ch III.6), a process $\{X_t\}$ is a Feller process if its semi-group of operators $\{P_t\}_{t\geq 0}$ satisfies the following two properties: (i) $P_t C_0 \subset C_0$ for all $t \geq 0$; (ii) for any $f \in C_0$ and $x \in E$, $P_t f(x) \to f(x)$ as $t \downarrow 0$. Feller processes have good path properties[1] and cover most processes we are interested in, e.g., the model (2.1).

In probability theory, Feller processes are more often characterized not using the transition function or semi-group of operators introduced above, but rather in terms of the infinitesimal operator. It is defined as follows: A function $f \in C_0$ is said to belong to the domain $D(\mathcal{A})$ of the infinitesimal operator $\mathcal{A}$ of a Feller process X if the following limit exists:

[1]By Rogers and Williams (2000, Ch III.7-9), the canonical Feller process always admits a Cadlag modification (the path of the process is right continuous and has left limits) and satisfies the strong Markov property

$$\mathcal{A}f = \lim_{t \downarrow 0} \frac{P_t f - f}{t} \tag{2.2}$$

with respect to the sup-norm of C_0.[2] Clearly, $\mathcal{A}$ is a linear operator from $D(\mathcal{A})$ to C_0. It can be seen from (2.2) immediately, that the following holds *P-a.s.* for $f \in D(\mathcal{A})$

$$E\left(\frac{f(X_{t+\Delta}) - f(X_t)}{\Delta}|_t\right) = \mathcal{A}f(X_t) + o(\Delta),$$

as $\Delta \downarrow 0$. In this sense, the infinitesimal operator indeed describes the expected movement of the process in an infinitesimally small time interval. In fact, it can be proved that the infinitesimal operator is equivalent to the semi-group of operators (and hence also the transition function) in fully characterizing the dynamics of a Feller process (see the Hill-Yoshida theorem in Dynkin (1965)).

For the model in (2.1), the infinitesimal operator always has a closed-form expression (see Kallenberg (2002, Thm 19.24) and Rogers and Williams (2000, Vol1, Thm III.13.3 and Vol2, Ch V.2)), given by

$$\mathcal{A}_\theta^D f(x) = \sum_{i=1}^d b_i(x;\theta) f_i'(x) + \frac{1}{2} \sum_{i,j=1}^d a_{i,j}(x;\theta) f_{i,j}''(x) \tag{2.3}$$

when X_t is a pure diffusion,

$$\begin{aligned} \mathcal{A}_\theta^P f(x) &= \sum_{i=1}^d b_i(x;\theta) f_i'(x) + \frac{1}{2} \sum_{i,j=1}^d a_{i,j}(x;\theta) f_{i,j}''(x) \\ &\quad + \lambda(x,\theta) \int \left[f(x+c) - f(x)\right] d\nu(c,\theta), \end{aligned} \tag{2.4}$$

when J_t is of Poisson-type, and

$$\mathcal{A}_\theta^L f(x) = \sum_{i=1}^d b_i(x;\theta) f_i'(x) + \frac{1}{2} \sum_{i,j=1}^d a_{i,j}(x;\theta) f_{i,j}''(x)$$

[2]Without using the sup-norm, Hansen and Scheinkman (1995) define infinitesimal operator in the Hilbert space $L^2(Q)$ where Q is an invariant (stationary) distribution of the process. This Hilbert space-based definition is needed in Hansen and Scheinkman (1995) for analyzing such properties as time reversibility, which is not needed in this paper.

$$+\int \left[f(x+c) - f(x) - f'(x)\, c 1_{0<|c|<1}\right] \pi(dc;\theta), \tag{2.5}$$

when J_t is a Levy process[3], where $f \in D(\mathcal{A})$, $x \in \mathbb{R}^d$, and

$$a_{ij}(x;\theta) = \sum_{k=1}^{d} \sigma_{i,k}(x;\theta)\sigma_{j,k}(x;\theta) \tag{2.6}$$

To illustrate the richness of information contained in the infinitesimal operator, we consider a univariate diffusion model defined as $dX_t = b(X_t)dt + \sigma(X_t)dW_t$. By (2.3), the infinitesimal operator for this univariate diffusion is

$$\mathcal{A}f(x) = b(x)f'(x) + \frac{1}{2}\sigma^2(x)f''(x)$$

Clearly the term involving the first derivative of the function $f(\cdot)$ is related to the dynamics of drift, and the term involving the second derivative to the dynamics of the diffusion function. This is consistent with the intuition that drift describes dynamics of the mean and diffusion describes that of the variance of the process (see Nelson (1990) for further discussion). Consider an infinitesimal change in this univariate diffusion process. By (2.3) and (2.6), for any $f \in D(\mathcal{A})$, it holds *P-a.s.* that

$$E\left(\frac{f(X_{t+\Delta}) - f(X_t)}{\Delta}|\mathfrak{F}_t\right) = b(X_t)f^{'}(X_t) + \frac{1}{2}\sigma^2(X_t)f^{''}(X_t) + o(\Delta), \tag{2.7}$$

as $\Delta \downarrow 0$. Therefore, the dynamics of $\{X_t\}$ are characterized completely by the drift and diffusion coefficients and hence by the infinitesimal operator. In fact, it follows from (2.7) that (Stanton, 1997):

$$b(X_t) = \lim_{\Delta\to 0} E\left[\frac{X_{t+\Delta} - X_t}{\Delta}|X_t\right], \sigma^2(X_t) = \lim_{\Delta\to 0} E\left[\frac{(X_{t+\Delta} - X_t)^2}{\Delta}|X_t\right]$$

which illustrates why $b(\cdot)$ and $\sigma^2(\cdot)$ are called instantaneous conditional mean and variance, respectively.

[3]Note that μ and σ_L^2, as components of Levy characteristics, do not appear explicitly here since they are combined into the drift and diffusion terms of the SDE (2.1). When only a Levy process is considered, the infinitesimal operator has both μ and σ_L^2 explicitly.

Since the Markov model in (2.1) is a Feller process, we now have three complete characterizations of the dynamics: transition function (or transition density) $P(t, x, \Gamma)$, semi-group of operators $\{P_t\}$, and infinitesimal operator $\mathcal{A}$. The transition density has already been used intensively in econometric inferences, not only in estimation (Lo 1988; Ait-Sahalia 2002; Yu, 2007) but also in hypothesis testing (Ait-Sahalia, Fan and Peng 2009; Hong and Li 2005). As is well known, the transition density of most continuous-time models has no closed form, and methods based on it are computationally burdensome and inconvenient.

In contrast, it is obvious from (2.3)-(2.6) that the infinitesimal operator is always analytic and fully characterizes the dynamics. This attractive property makes the infinitesimal operator a convenient tool for econometric inferences. It has been used in identification (Hansen, Scheinkman and Touzi 1998), estimation (Hansen and Scheinkman 1995; Kessler and Sorenson 1999; Duffie and Glynn, 2004) and also hypothesis testing (Kanaya 2007; Song 2011). However, the cited methods of generating moment conditions using the infinitesimal operator have major drawbacks, such as identification up to scale (discussed in Section 2.1.4). In the following, I shall derive alternative and convenient infinitesimal operator-based conditional moment restrictions by a new technique, which can identify all parameters uniquely.

2.1.2 Conditional Moment Restrictions

To obtain moment conditions by utilizing the closed-form infinitesimal operator, I consider a transformation based on the celebrated "martingale problems".

Following Ch. 5.4 of Karatzas and Shreve (1991), a probability measure P on $\left(C[0,\infty)^d, \mathcal{B}(C[0,\infty)^d)\right)$, under which

$$M_t^f = f(X_t) - f(X_0) - \int_0^t (\mathcal{A}f)(X_s)ds \tag{2.8}$$

is a martingale for every $f \in D(\mathcal{A})$, is called a solution to the martingale problem associated with the operator $\mathcal{A}$.

How are the "martingale problems" related to Model (2.1)? It is well-known that a SDE has two types of solutions: strong solutions and weak solutions (see Karatzas and Shreve (1991, Ch 5.2-3) or Rogers and Williams (2000, Ch V.2-3) for details). When the drift and diffusion terms of a SDE satisfy the Lipschitz and linear growth conditions (Protter, 2005), there exists a strong solution to the SDE. However, for general drift, diffusion, and jump terms, a strong solution may not exist; in this case, probabilists usually attempt to solve the SDE in the "weak" sense of finding a solution with the right probability law. The martingale problem is a variation of this "weak solution approach" developed by Strook and Varadhan (1969) and is in fact equivalent to the weak solution of a SDE. That is, the process $\{X_t\}$ is a weak solution to the SDE (2.1) if and only if

$$M_t^f(\theta) = f(X_t) - f(X_0) - \int_0^t (\mathcal{A}_\theta f)(X_s)ds \tag{2.9}$$

is a martingale for every $f \in D(\mathcal{A})$, where $\mathcal{A}_\theta$ is defined as in (2.3)-(2.5). For detailed discussion and proof, see Ch V.19-20 of Rogers and Williams (2000), Theorem 21.7 of Kallengberg (2002), or Proposition 2.4 of Ch VII in Revuz and Yor (2005).

Loosely speaking, for a strong solution only $\{X_t\}$ is constructed with respect to a filtration generated by a given Brownian motion $\{W_t\}$, while in the case of a weak solution, not only $\{X_t\}$ but also the driving Brownian motion are built as parts of the solution. The difference between strong and weak solutions, is intuitively very similar to that between a random variable and its distribution. Since econometric inferences are only concerned with the dynamic probability laws of the process on the space of trajectories instead of with specific sample paths, it is sufficient to consider a weak solution to the SDE (2.1) here.

We have now shown that the identification of the multivariate time-homogeneous continuous-time Markov process in (2.1) is equivalent to the martingale property of the transformed processes in (2.9). By the uniqueness of the solution to (2.1) in Assumption D.1, the martingale property can be written as a conditional moment restriction:

$$E\left[M_t^f(\theta_0)|\mathcal{I}_{t'}\right] = M_{t'}^f(\theta_0)$$

for a unique $\theta_0 \in \Theta$ and any $f \in D(\mathcal{A})$ and $t' < t$, where $\text{callI}_{t'} = \sigma\{X_{t''}\}_{t''<t'}$ is the sigma-field generated by the past information of $\{X_t\}$ at time t'. For convenience, I state the following equivalent conditional moment restrictions using the martingale difference sequence (*m.d.s.*) property for the first-order difference of the transformed process $M_t^f(\theta)$:

$$E\left[Z_t^f(\theta_0)|\mathcal{I}_{t'}\right] = 0 \tag{2.10}$$

for a unique $\theta_0 \in \Theta$ and any $f \in D(\mathcal{A})$, $t' < t$ and $\Delta > 0$, where with $Z_t^f(\theta) = M_t^f(\theta) - M_{t-\Delta}^f(\theta)$ and $\mathcal{I}_{t'} = \sigma\{X_{t''}\}_{t''<t'}$. By the Markov property of the process X_t, (2.10) is equivalent to

$$E\left[Z_t^f(\theta_0)|X_{t-\Delta}\right] = 0 \tag{2.11}$$

for a unique $\theta_0 \in \Theta$ and any $f \in D(\mathcal{A})$ and $\Delta > 0$.

2.1.3 Choice of Test Functions

Observe that in (2.11) we have infinitely many conditional moment restrictions because there are usually an infinite number of test functions $f(\cdot)$ in the domain $D(\mathcal{A})$. Due to the difficulty of exhausting all possible functional forms in $D(\mathcal{A})$, it is very burdensome in practice (although maybe not impossible) to construct an estimator based on these infinitely many conditional moment conditions. This is a general problem which appears not only in my study but also for all other papers employing infinitesimal operators, such as Hansen and Scheinkman (1995), Conley, Hansen, Luttmer and Scheinkman (1997), and Kanaya (2007). To tackle such a difficulty, the space of test functions has to be reduced to an equivalent subclass in such a way that no identification information is lost.

In the following, I discuss choices of test functions for three popular types of model (2.1), i.e., diffusion, jump-diffusion, and Levy jump models. with infinitesimal operators as defined in (2.3)- (2.5). For pure diffusions, a celebrated theorem in probability theory allows the construction of a subclass of $D(\mathcal{A})$ for the martingale characterization (2.9) without losing identification information. This subclass consists of only finitely many function forms. By Proposition 4.6 and Remark 4.12 of Karatzas and Shreve (1991, Chp. 5.4), the process $\{X_t\}$ is a weak solution to the SDE in (2.1) without jump terms if[4] it satisfies the martingale problem with $\mathcal{A}$ as the infinitesimal operator of $\{X_t\}$ for the choices $f(x) = x_i$ and $f(x) = x_i x_j$ with $1 \leq i, j \leq d$. Therefore, the conditional moment restric-

[4]Under fairly general conditions, the converse of this result only holds with local martingale replacing martingale. The difference between local martingale and martingale in continuous-time finance matters mainly for asset price bubbles; see Jarrow, Protter, and Shimbo (2006, 2010) for models of asset price bubbles in a continuous-time local martingale framework. Here I assume that the diffusion functions $\sigma_{i,k}(\cdot)$: $\mathbb{R}^d \to \mathbb{R}$ are continuous functions and hence bounded on compact sets. By Karatzes and Shreve (1991, Propostion 4.11), the weak solution is equivalent to the martingale property now.

tion (2.11) for pure diffusion models has $Z_t^f(\theta)$ as a vector with components for $i, j = 1, \cdots, d$

$$\begin{aligned}
Z_t^i(\theta_0) &= M_t^{x_i}(\theta_0) - M_{t-\Delta}^{x_i}(\theta_0) = X_t^i - X_{t-\Delta}^i - \int_{t-\Delta}^t b_i(X_s;\theta_0)ds \\
Z_t^{i,i}(\theta_0) &= M_t^{x_i x_i}(\theta_0) - M_{t-\Delta}^{x_i x_i}(\theta_0) \\
&= \left(X_t^i\right)^2 - \left(X_{t-\Delta}^i\right)^2 - \int_{t-\Delta}^t \left[2b_i(X_s;\theta_0)X_s^i + \sum_{k=1}^d \sigma_{i,k}(X_s;\theta_0)^2\right] ds \\
Z_t^{i,j}(\theta_0) &= M_t^{x_i x_j}(\theta_0) - M_{t-\Delta}^{x_i x_j}(\theta_0) \\
&= X_t^i X_t^j - X_{t-\Delta}^i X_{t-\Delta}^j - \int_{t-\Delta}^t \Big[b_i(X_s;\theta_0)X_s^j + b_j(X_s;\theta_0)X_s^i \\
&\quad + \frac{1}{2}\sum_{k=1}^d \sigma_{i,k}(X_s;\theta_0)\sigma_{j,k}(X_s;\theta_0)\Big] ds \qquad (2.12)
\end{aligned}$$

for $i \neq j$ To gain better understanding of the characterization (2.12), we consider the simplified version for univariate diffusion models:

$$\begin{aligned}
M_t^x(\theta_0) &= X_t - X_0 - \int_0^t b(X_s;\theta_0)ds \\
M_t^{x^2}(\theta_0) &= X_t^2 - X_0^2 - \int_0^t \left[2b(X_s;\theta_0)X_s + \sigma^2(X_s;\theta_0)\right] ds
\end{aligned}$$

are both martingales A simple but important example can give us more insight.

Example 2.1.1: Levy Characterization of Brownian Motion.

Suppose $b(\cdot) \equiv 0$ and $\sigma(\cdot) \equiv 1$ for the univariate diffusion models $dX_t = b(X_t)dt + \sigma(X_t)dW_t$ with $X_0 = 0$. Then clearly the solution $X_t = W_t$ is the standard Brownian motion. Plugging the drift and diffusion terms into M_t^x and $M_t^{x^2}$ above, we now have

$$\begin{aligned}
M_t^x &= X_t \\
M_t^{x^2} &= X_t^2 - t
\end{aligned}$$

are both martingales. That is, according to the infinitesimal operator-based martingale characterization employed in this paper, X_t is a standard

Brownian motion if and only if both X_t and $X_t^2 - t$ are martingales. This is exactly the Levy Characterization Theorem (Øksendal, 2003, Theorem 8.6.1) which gives necessary and sufficient conditions for characterizing the standard Brownian Motion. This example shows that the infinitesimal operator-based characterization we depend on is an extension of this result to general diffusion models.

However, for jump-diffusion and Levy jump models with infinitesimal operators defined in (2.4) and (2.5) respectively, there does not exist such simplified functional forms to reduce the space $D(\mathcal{A})$. Based on Kanaya (2007) who chooses exponential functions via the basis of $D(\mathcal{A})$, I here propose to use the exponential function $\exp\left[-\left(x_1^2 + \cdots + x_d^2\right)/2\right]$. Therefore, the conditional moment restriction (2.11) has the form

$$Z_t^f(\theta) = e^{-\left(X_{1,t}^2+\cdots+X_{d,t}^2\right)/2} - e^{-\left(X_{1,t-\Delta}^2+\cdots+X_{d,t-\Delta}^2\right)/2} - \int_{t-\Delta}^{t} \mathcal{A}_\theta e^{-\left(X_{1,s}^2+\cdots+X_{d,s}^2\right)/2} ds \tag{2.13}$$

where $\mathcal{A}_\theta \exp\left[-\left(X_{1,s}^2 + \cdots + X_{d,s}^2\right)/2\right]$ is equal to

$$\begin{aligned} & \mathcal{A}_\theta^P e^{-\left(X_{1,s}^2+\cdots+X_{d,s}^2\right)/2} \\ = \; & e^{-\left(X_{1,s}^2+\cdots+X_{d,s}^2\right)/2} \left\{ -\sum_{i=1}^{d} b_i(X_s;\theta) X_{i,s} + \frac{1}{2}\sum_{i,j=1}^{d} a_{i,j}(X_s;\theta) X_{i,s} X_{j,s} \right. \\ & \left. -\frac{1}{2}\sum_{i,j=1}^{d} a_{i,j}(X_s;\theta) + \lambda(X_s,\theta) \int \left[e^{-c\cdot X_s - |c|^2/2} - 1 \right] d\nu\,(c,\theta) \right\}, \end{aligned} \tag{2.14}$$

when J_t is the compound Poisson process J_t^P, and

$$\begin{aligned} & \mathcal{A}_\theta^L e^{-\left(X_{1,s}^2+\cdots+X_{d,s}^2\right)/2} \\ = \; & e^{-\left(X_{1,s}^2+\cdots+X_{d,s}^2\right)/2} \left\{ -\sum_{i=1}^{d} b_i(X_s;\theta) X_{i,s} + \frac{1}{2}\sum_{i,j=1}^{d} a_{i,j}(X_s;\theta) X_{i,s} X_{j,s} \right. \\ & \left. -\frac{1}{2}\sum_{i,j=1}^{d} a_{i,j}(X_s;\theta) + \int \left[e^{-c\cdot X_s - |c|^2/2} - 1 + c \cdot X_s 1_{0<|c|<1} \right] \pi\,(dc,\theta) \right\} \end{aligned} \tag{2.15}$$

when J_t is the Levy process J_t^L.

Along with (2.12)-(2.14), the resulting characterization (2.11) greatly simplifies the conditional moment restrictions. It can be observed that the conditional moment restrictions are expressed explicitly in terms of drift and diffusion terms and can be used directly, in contrast to transition density-based methods like Lo (1988) and Ait-Sahalia (2002, 2008) which must either approximate the transition density or solve for it numerically since it rarely has a closed-form. The simplified conditional moment restrictions are particularly convenient for multivariate models for which transition density methods are extremely complicated and computationally inconvenient.

2.1.4 Comparison with HS Moment Conditions for Diffusion Models

In this section, my infinitesimal operator-based conditional moment restrictions derived above will be compared to **HS** moment conditions. To show the relative merits, consider a univariate diffusion model

$$dX_t = b\left(X_t; \kappa\right) dt + \sigma\left(X_t; \gamma\right) dW_t \tag{2.16}$$

where drift and diffusion parameters are denoted separately as $\kappa \in \mathbb{R}$ and $\gamma \in \mathbb{R}$. **HS** moment conditions are in the form of expectations of the infinitesimal operator $\mathcal{A}_{\theta,\gamma}$ for (2.16), one unconditional (C1) and the other conditional (C2):

$$\begin{aligned} C1 \quad : \quad 0 &= E\left[\mathcal{A}_{\kappa_0,\gamma_0} \cdot \psi(X_0, \kappa_0, \gamma_0)\right] \\ &= E\left[b(X_0; \kappa_0)\frac{\partial \psi(X_0, \kappa_0, \gamma_0)}{\partial X_0} + \frac{1}{2}\sigma^2(X_0; \gamma_0)\frac{\partial^2 \psi(X_0, \kappa_0, \gamma_0)}{\partial X_0^2}\right] \end{aligned} \tag{2.17}$$

where κ_0 and γ_0 are the true parameter values and $\psi(\cdot, \kappa, \gamma)$ is a sufficiently differentiable function in $D\left(\mathcal{A}_{\kappa,\gamma}\right)$, the domain of the infinitesimal operator $\mathcal{A}_{\kappa,\gamma}$. The assumed stationarity of $\{X_t\}$ implies that the unconditional expectation $E_{X_t}\left[\psi\left(X_t; \kappa, \gamma\right)\right]$ does not depend on the time t; thus, $\frac{\partial}{\partial t} E_{X_t}\left[\psi\left(X_t; \kappa, \gamma\right)\right] = 0$, which yields (2.17).

Two functions ψ_0 and ψ_1 satisfying smoothness and regularity conditions are taken by **HS** to form the "back to the future" C2 moment condition:

$$\begin{aligned}
& E\left\{\left[\mathcal{A}_{\kappa_0,\gamma_0} \cdot \psi_1(X_1, \kappa_0, \gamma_0)\right] \times \psi_0(X_0, \kappa_0, \gamma_0) - \left[\mathcal{A}^*_{\kappa,\gamma} \cdot \psi_0(X_0, \kappa_0, \gamma_0)\right] \times \psi_1(X_1, \kappa_0, \gamma_0)\right\} \\
= \; & E\left\{\left[b(X_1; \kappa_0)\frac{\partial \psi_1(X_1, \kappa_0, \gamma_0)}{\partial X_1} + \frac{1}{2}\sigma^2(X_1; \gamma_0)\frac{\partial^2 \psi_1(X_1, \kappa_0, \gamma_0)}{\partial X_1^2}\right] \times \psi_0(X_0, \kappa_0, \gamma_0)\right\} \\
& -E\left\{\left[b(X_0; \kappa_0)\frac{\partial \psi_0(X_0, \kappa_0, \gamma_0)}{\partial X_0} + \frac{1}{2}\sigma^2(X_0; \gamma_0)\frac{\partial^2 \psi_0(X_0, \kappa_0, \gamma_0)}{\partial X_0^2}\right] \times \psi_1(X_1, \kappa_0, \gamma_0)\right\} \\
= \; & 0 \qquad (2.18)
\end{aligned}$$

where $\mathcal{A}^*_{\kappa,\gamma}$ is the infinitesimal operator associated with the time-reversed process of $\{X_t\}$. In this case, $\mathcal{A}^*_{\kappa,\gamma}$ is equal to $\mathcal{A}_{\kappa,\gamma}$ since univariate stationary diffusions are time reversible (Kent, 1978) under regularity conditions. Similar to (2.17), Eqnarray (2.18) follows from the assumed stationarity of $\{X_t\}$, which implies that $E_{X_t, X_{t+\Delta}}\left[\psi_0\left(X_t; \kappa, \gamma\right)\psi_1\left(X_{t+\Delta}; \kappa, \gamma\right)\right]$ does not depend on time t and hence $(\partial/\partial t)\, E_{X_t, X_{t+\Delta}}\left[\psi_0\left(X_t; \kappa, \gamma\right)\psi_1\left(X_{t+\Delta}; \kappa, \gamma\right)\right] = 0$.

My infinitesimal operator-based conditional moment condition for the case (2.16) is $E\left[Z_t(\kappa_0, \gamma_0)|X_{t-\Delta}\right] = 0$, where $Z_t(\kappa_0, \gamma_0) = \left(Z^x_t(\kappa_0, \gamma_0), Z^{x^2}_t(\kappa_0, \gamma_0)\right)'$ and

$$\begin{aligned}
Z^x_t(\kappa_0, \gamma_0) &= M^x_t(\kappa_0, \gamma_0) - M^x_{t-\Delta}(\kappa_0, \gamma_0) \\
&= X_t - X_{t-\Delta} - \int_{t-\Delta}^t b(X_s; \kappa_0)ds \\
Z^{x^2}_t(\kappa_0, \gamma_0) &= M^{x^2}_t(\kappa_0, \gamma_0) - M^{x^2}_{t-\Delta}(\kappa_0, \gamma_0) \\
&= X_t^2 - X_{t-\Delta}^2 - \int_{t-\Delta}^t \left[2b(X_s; \kappa_0)X_s + \sigma^2(X_s; \gamma_0)\right]ds \qquad (2.19)
\end{aligned}$$

As pointed out by Ait-Sahalia (2007) and Ait-Sahalia and Mykland (2008), it can be seen from (2.17) and (2.18) that no "natural" conditional moments of the process are available to act as the **HS** moment conditions since explicit expressions for conditional mean, variance, skewness and so on are not in closed-form. Moreover, both C1 and C2 are in the form of the infinitesimal operator applied to arbitrary test functions. Additional work is needed to choose suitable test functions, as in Ait-Sahalia and Mykland (2008) who rigorously analyze the impact of different test functions on the variance of **HS** GMM estimators by deriving the closed-form expansion of the estimator's asymptotic distribution. In contrast, my infinitesimal operator-based conditional moment condition is derived by a "natural" choice of test functions which preserves the advantageous property of complete identification and features an intuitive interpretation in terms of instantaneous conditional mean and variance.

More importantly, as discussed in Ait-Sahalia and Mykland (2008), the **HS** moment conditions in (2.17) and (2.18) do not make efficient use of the entire range of information contained in the sample and hence cannot have full identification of all parameters. To see this serious problem, we multiply the drift and diffusion functions by the constants a and $\sqrt{a}$, respectively. This results in identical moment conditions for both C1 and C2 with a different set of parameters. Hence the parameters cannot be identified uniquely in the **HS** moment conditions; they are only identified up to scale. However, if we do the same multiplication for the conditional moment conditions (2.19), the resulting moment conditions are now:

$$E\left[Z_t(\kappa_0, \gamma_0, \lambda)|X_{t-\Delta}\right] = 0 \tag{2.20}$$

for any $t' < t$, where $Z_t(\kappa_0, \gamma_0, \lambda) = \left(Z_t^x(\kappa_0, \gamma_0, \lambda), Z_t^{x^2}(\kappa_0, \gamma_0, \lambda)\right)'$ and

$$\begin{aligned}
Z_t^x(\kappa_0, \gamma_0, \lambda) &= M_t^x(\kappa_0, \gamma_0, \lambda) - M_{t-\Delta}^x(\kappa_0, \gamma_0, \lambda) \\
&= \frac{X_t - X_{t-\Delta}}{\lambda} - \int_{t-\Delta}^{t} b(X_s; \kappa_0) ds \\
Z_t^{x^2}(\kappa_0, \gamma_0, \lambda) &= M_t^{x^2}(\kappa_0, \gamma_0, \lambda) - M_{t-\Delta}^{x^2}(\kappa_0, \gamma_0, \lambda) \\
&= \frac{X_t^2 - X_{t-\Delta}^2}{\lambda} - \int_{t-\Delta}^{t} \left[2b(X_s; \kappa_0)X_s + \sigma^2(X_s; \gamma_0)\right] ds
\end{aligned}$$

(2.20) is clearly a different moment condition than the original one before the multiplication. Therefore, all the parameters are uniquely identified by the proposed conditional moment restriction (2.19). Intuitively, the improvement of my conditional moment conditions over **HS** moment conditions is due to the unique characterization of the process dynamics by the "martingale problems" in (2.9).

A specific example illustrates the problem more concretely. Consider the stationary Ornstein-Uhlenbeck process $dX_t = -\kappa X_t dt + \sigma dW_t$ with $\kappa > 0$. We follow Conley et al. (1997) to choose the score vector of the model-implied stationary density as the test function $\psi(\cdot)$. For this special example, the stationary density is

$$g\left(x, \kappa, \sigma^2\right) = \frac{1}{\sqrt{2\pi}\,(\sigma^2/(2\kappa))} \exp\left[-\frac{x^2}{2\sigma^2/(2\kappa)}\right] \tag{2.21}$$

which is a normal density with mean 0 and variance $\sigma^2/(2\kappa)$. Then by taking derivatives and manipulating the terms, it can be shown that

$$\psi = \begin{pmatrix} \psi_1 \\ \psi_2 \end{pmatrix} = \begin{pmatrix} \frac{1}{\kappa} - \frac{x^2}{\sigma^2} \\ -\frac{1}{\sigma^2} + \frac{\kappa x^2}{(\sigma^2)^2} \end{pmatrix}$$

, which is a two-dimensional vector. The **HS** C1 moment condition is[5], by (2,17)

[5]The results with the C2 moment condition are very similar to those with C1 moment condition.

and (2.21),

$$E\left[-\kappa X_t \frac{\partial \psi_1}{\partial x}|_{x=X_t} + \frac{1}{2}\sigma^2 \frac{\partial^2 \psi_1}{\partial x^2}|_{x=X_t}\right] = E\left[\frac{2\kappa}{\sigma^2}X_t^2 - 1\right] = 0 \tag{2.22}$$

for $\psi_1$and

$$E\left[-\kappa X_t \frac{\partial \psi_2}{\partial x}|_{x=X_t} + \frac{1}{2}\sigma^2 \frac{\partial^2 \psi_2}{\partial x^2}|_{x=X_t}\right] = E\left[\frac{2\kappa}{\sigma^2}X_t^2 - 1\right] = 0 \tag{2.23}$$

for ψ_2. Observe that (2.22) and (2.23) are actually the same and hence we only have one moment condition:

$$E\left[\frac{2\kappa}{\sigma^2}X_t^2 - 1\right] = 0 \tag{2.24}$$

Obviously, only κ/σ^2 can be estimated based on **HS** C1 moment conditions. This confirms the conclusion above that parameters are only identified up to scale in **HS** moment conditions. For my infinitesimal operator based conditional moment restriction, it follows from (2.20) that

$$E\left[\left(\begin{array}{c} X_t - X_{t-\Delta} - \int_{t-\Delta}^{t} -\kappa X_s ds \\ X_t^2 - X_{t-\Delta}^2 - \int_{t-\Delta}^{t}\left[-2\kappa X_s^2 + \sigma^2\right] ds \end{array}\right)|X_{t-\Delta}\right] = 0 \tag{2.25}$$

which, as a two dimensional moment condition, does identify both σ^2 and κ.

2.2 Local Empirical Likelihood Estimation

2.2.1 Conditional Moment Restrictions and Asymptotic Schemes

As shown in Section 2, the identification of the model (2.1) is equivalent to the conditional moment restriction in (2.11): there exists a unique θ_0 such that $E\left[Z_t^f(\theta_0)|X_{t-\Delta}\right] = 0$ for any $\Delta > 0$ with choices of $f(\cdot)$ as in (2.12)-(2.15). It can be observed that to characterize the process dynamics at every time point, the conditional moment restriction has to be satisfied for each possible Δ consistent with the continuous-time framework. This is similar to the case of transition density-based methods: we must consider the transition density $p\left(X_t|X_{t-\Delta};\theta\right)$ for each $\Delta > 0$ in studying the dynamics of the process.

However, the processes can only be observed discretely in time with $\{X_{\tau\Delta}\}_{\tau=1}^{n}$ over a time span T, the sampling interval Δ, and the sample size $n = T/\Delta$. Once Δ is fixed, information about the process dynamics inside this sampling interval is lost[6]. But since the models considered are parametric models, the parameters can still be estimated consistently using the conditional moment restriction in (2.11) or transition density $p\left(X_t|X_{t-\Delta};\theta\right)$ for a given Δ.

As discussed earlier, the transition density $p\left(X_t|X_{t-\Delta};\theta\right)$ is rarely available in closed form, and this poses an impediment to implementing the MLE. Therefore, we shall depend on the conditional moment restriction, $E\left[Z_t^f(\theta)|X_{t-\Delta}\right] = 0$ for some fixed Δ, to construct an estimator. For the identification of the model parameters in (2.1), we can, based on the discussions above and Assumption 2.2.1, impose the following identification condition:

Assumption 2.2.1: There exists a unique $\theta_0 \in \Theta$ such that $E\left[Z_t^f(\theta)|X_{t-\Delta}\right] = 0$ holds for some fixed Δ.

Note that $Z_t^f(\theta)$ involves integrals of the form $\int_{(\tau-1)\Delta}^{\tau\Delta} g(X_s)ds$. To compute them, I assume that M observations exist in each sampling interval Δ with $M \to \infty$. Then we can approximate these integrals by[7] $\int_{(\tau-1)\Delta}^{\tau\Delta} g(X_s)ds =$

[6]In fact, this is the "aliasing problem" in Phillips (1973) about the identification of continuous-time econometric models

[7]Another method is to generate multiple high-frequency sample paths in the sampling interval Δ using Euler discretization schemes. Then integrals can be approximated by taking the

$\frac{\Delta}{M+1}\sum_{m=(\tau-1)\Delta+\frac{\Delta}{M}}^{\tau\Delta} g\left(X_m\right)+O_{a.s.}\left(M^{-2}\right)$As a result, the asymptotic schemes employed here are $n = T/\Delta \to \infty$ from the standard discrete-time econometrics and the "in-fill" represented by $\Delta/M \to 0$ specific to the continuous-time econometrics[8]. In contrast, both **HS** and Ait-Sahalia (2002, 2008) only need $n \to \infty$. Hence their methods are more applicable than the proposed LEL estimator in terms of types of data. This is the cost we pay by switching from the transition density to the convenient infinitesimal operator and avoiding numerical and simulation-based procedures. As a benefit, we obtain a convenient estimation method for general multivariate Markov models in a unified framework[9].

The "in-fill" asymptotic scheme is very different from estimation methods based on the Euler discretization of continuous-time models, although both lead to shrinking sampling intervals. The former is only due to a numerical integral while the latter discretizes the stochastic model. I conduct simulation studies of the proposed LEL estimator with comparisons to the estimators via the discretized versions of the models using Euler discretization schemes. Results show that the proposed LEL estimator outperforms the Euler approximation schemes in situations relevant for financial models; see Section 4 for details.

average of the generated sample paths. This method is similar to the simulated MLE approach in Brandt and Santa-Clara (2002) and incurs large computational burdens due to simulations.

[8]Bandi and Phillips (2003) argue that both $\Delta \to 0$ and $T \to \infty$ are needed to estimate continuous-time (diffusion) processes fully non-parametrically. Specifically, $T \to \infty$ is needed for nonparametrically estimating the drfit function. In contrast, we only consider parametric models here and hence $T \to \infty$ is not necessary.

[9]Note that when M is fixed, e.g., $M = 1$, the "in-fill" approximation above becomes

$$\int_{(\tau-1)\Delta}^{\tau\Delta} g(X_s)ds = \frac{\Delta}{2}\left[g\left(X_{\tau\Delta}\right)+g\left(X_{(\tau-1)\Delta}\right)\right]+O_P\left(\Delta^2\right)$$

where $\Delta \to 0$. This is in fact the approximation scheme adopted in Pan (2002) and Hong, Lee and Song (2009). The performance of this approximation depends on the speed of mean-reversion of the process and the size of Δ. Pan (2002), Song (2011) and Hong, Lee and Song (2009) find that for daily and even monthly data, such a simple approximation works reasonably well in finite samples. In the simulation studies of Section 2.3, I also find similar good finite sample performances for jump-diffusion and Levy jump-diffusion models at the daily frequency.

2.2.2 The Estimator

The model framework I consider is in fact a bit more general:

$$E\left[u(X_t, X_{t-\Delta}; \theta) | X_{t-\Delta}, \cdots, X_{t-m\Delta}\right] = 0 \tag{2.26}$$

where $\{X_t\}_{t=\Delta,2\Delta,\cdots}$ is a strictly stationary[10] and m-th order Markov process. Obviously, the identification eqnarray (2.11) with a fixed Δ for our continuous-time Markov model (2.1) is a special case with m=1 and $u(\cdot)$ containing integrals of the form $\int_{(\tau-1)\Delta}^{\tau\Delta} g(X_s)ds$ as in the last section.

Suppose $x_t = (X_{t-\Delta}, \cdots, X_{t-m\Delta})'$ and $y_t = (X_t, x_t')'$; then (2.26) can be restated as:

$$E\left[u(y_t; \theta) | x_t\right] = 0.$$

where the dimensions of the vectors are $dm \times 1$ for x_t, $(d+1)m \times 1$ for y_t, and $q \times 1$ for the function $u(\cdot; \theta)$. Such a framework is of independent interest beyond the continuous-time settings in the current paper. In fact, it covers many popular models in discrete time such as a simple AR(1) model with m.d.s. (Gospodinov and Otsu, 2009) and has been a focus in econometrics for many decades; see Dominguez and Lobato (2004), Donald, Imbens and Newey (2003), **KTA**, Smith (2007), Carrasco and Florens (2008), and Tripathi and Kitamura (2003) for references.

Let $D(x) = E\left[\nabla_\theta u(y_t, \theta) | x_t = x\right]$ and $V(x) = E\left[u(y_t, \theta) u(y_t, \theta)' | x_t = x\right]$ where ∇_θ is the gradient operator indexed by the parameter θ. For I.I.D data, an efficient estimator advocated by Chamberlain (1987), Robinson (1987), and Newey (1990)

[10] It is worth pointing out that the stationarity assumption is imposed only for proving asymptotic properties of the proposed LEL estimator. The derived identification condition (2.11), which is a special case of (3.1), does not depend on the stationarity assumption and hence is potentially useful for studying nonstationary models considered in Bandi and Phillips (2003) and Bandi and Reno (2008).

can be constructed by estimating the optimal instrument $a^*(x) = D(x)' V^{-1}(x)$ in the first step using a preliminary estimator $\widetilde{\theta}$, and then implementing the optimal GMM by the estimated optimal instrument $\widetilde{a}^*(x)$. The resultant estimator can be shown to achieve the semiparametric efficiency bound

$$I^{-1} = \left\{E\left[D(x)' V^{-1}(x) D(x)\right]\right\}^{-1} \tag{2.27}$$

provided in Chamberlian (1987). However, it is shown by Dominguez and Lobato (2004) that even if the optimal IV a^* were known, the moment condition $E[a^*(x_t) u(y_t;\theta)] = 0$ by which the estimator above is constructed may fail to identify θ, although the original model (2.26) succeeds. It is therefore important to impose (2.26) directly when estimating θ.

In recent years, several estimators have been proposed to directly estimate model (2.26) free of the identification problem including, for example the LEL method in **KTA**, method of forming unconditional moment restrictions based on approximating functions in Donald, Imbens and Newey (2003), and the minimum-distance estimator in Domínguez and Lobato (2004). Smith (2007) extends **KTA** using the Cressie-Read power divergence family of discrepancies, which includes the Local GMM in Gospodinov and Otsu (2009) as a special case.

However, most of this research is focused on the I.I.D. environment and cannot be employed directly in our case. In this section, I shall extend the LEL estimator in **KTA** to the time-series framework in (2.26). Specifically, I show that the LEL estimator is consistent, asymptotically normal and more importantly attains the semi-parametric efficiency bound provided in Carrasco and Florens (2008) for time series data. This asymptotic efficiency result is new in the literature and is of independent interest as an extension of results for I.I.D data in Newey (1990, 1993) and Kitamura, Tripathi and Ahn (2004). Such an extension is also important for the following additional reasons. First, as discussed earlier, the LEL approach employs the conditional moment restriction directly and hence is able to identify the parameter globally. Second, theoretical and simulation studies in the literature on unconditional moment condition models reveal that empirical likelihood methods enjoy better finite sample performance (Newey and Smith, 2004; Anatolyev, 2005). It is expected that LEL for the conditional moment restriction model (2.26) may also perform well in finite samples, especially when the dimension of moments grows with the sample size. Last, as discussed in Smith (2007) and proved in the next section, the LEL approach avoids explicit estimation of the conditional Jacobian and Hessian matrices $D(x)$ and $V(x)$ while achieving the semi-parametric efficiency bound.

Several alternative estimators also free of the identification problem are proposed in the literature for the conditional moment restrictions in time series. For example, Carrasco, Chernov, Ghysels and Florens (2007) extend the estimator based on a continuum of moment conditions in Carrasco and Florens (2000) from the I.I.D. to the time-series framework. A user-chosen number must be employed for the inversion of the covariance operator. Gospodinov and Otsu (2009) consider an m-th order Markov framework similar to our (2.26) here and propose a Local GMM estimator which is a special case of Smith (2007) and which facilitates the analysis of the bias reduction property using higher-order expansions. The LEL approach employed here, by contrast, is more appealing in the sense that natural empirical local conditional probabilities can be provided (Smith, 2007).

To construct the LEL estimator, we first define the positive weights

$$w_{ij} = \frac{\mathcal{K}\left[\left(x_i - x_j\right)/b_n\right]}{\sum_{j=1}^{n} \mathcal{K}\left[\left(x_i - x_j\right)/b_n\right]} \triangleq \frac{\mathcal{K}_{ij}}{\sum_{j=1}^{n} \mathcal{K}_{ij}} \tag{2.28}$$

where $\mathcal{K}(\cdot)$ is a kernel function and $b_n \in \mathbb{R}$ is a sequence of positive bandwidth numbers. Note that $\sum_{j=1}^{n} w_{ij} = 1$ is automatically satisfied. Let p_{ij} be the probability mass placed at $\left(x_i, y_j\right)$ by a discrete distribution supported on $\{x_1, \cdots, x_n\} \times \{y_1, \cdots, y_n\}$, which can be regarded as an estimate of the conditional probability $P\left\{y = y_j | x = x_i\right\}$. Then we can form the following maximization problem, which is essentially a "nonparametric maximum likelihood" approach:

$$\begin{aligned} &\max_{p_{ij}} \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} \log p_{ij} \\ s.t. p_{ij} &\geq 0, \sum_{j=1}^{n} p_{ij} = 1, and \sum_{j=1}^{n} u\left(y_j; \theta\right) p_{ij} = 0 \end{aligned} \tag{2.29}$$

for $i, j = 1, \cdots, n$.

To better understand (2.29), we consider the general maximization problem in the following form with the notation of the local Cressie-Read discrepancy criterion in Smith (2007):

$$\begin{aligned} &\min_{p_{ij}} \frac{1}{\gamma(\gamma+1)} \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} \left[\left(\frac{p_{ij}}{w_{ij}}\right)^{\gamma+1} - 1\right] \\ s.t. p_{ij} &\geq 0, \sum_{j=1}^{n} p_{ij} = 1, and \sum_{j=1}^{n} u\left(y_j; \theta\right) p_{ij} = 0 \end{aligned} \tag{2.30}$$

for $i, j = 1, \cdots, n$. When $\gamma = -1$, (2.30) reduces to (2.29). Therefore, the optimization problem (2.29) can be regarded as minimizing a distance defined by a special local Cressie-Read discrepancy criterion: That is, the distance between the conditional probabilities p_{ij} $(i, j = 1, \cdots, n)$ (incorporating the conditional moment restrictions through $\sum_{j=1}^{n} u\left(y_j; \theta\right) p_{ij} = 0$) and kernel weights w_{ij} $(i, j = 1, \cdots, n)$ (determined by data directly). Other cases of (2.30) with different values of γ will lead to local versions of different empirical likelihood estimators for unconditional moment conditions, such as the local exponential tiltering estimator in Kitamura and Stutzer (1997). Although I only focus on the LEL estimator in (2.29) here, the theory developed in this paper can actually be extended to the whole family in (2.30) by modifying the imposed regularity conditions suitably.

The problem (2.29) can be conveniently solved by a Lagrange multiplier method. Let

$$\mathcal{L}(\theta) = \sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}\log p_{ij} - \sum_{i=1}^{n}\mu_i\left(\sum_{j=1}^{n} p_{ij} - 1\right) - \sum_{i=1}^{n}\lambda_i'\left(\sum_{j=1}^{n} u\left(y_j;\theta\right)p_{ij} = 0\right) \quad (2.31)$$

where $\mu_1,\cdots,\mu_n$ and $\lambda_1,\cdots,\lambda_n$ are the Lagrange multipliers for the second and third sets of constraints respectively. By **KTA**, the solution is

$$\widehat{p}_{ij} = \frac{w_{ij}}{1+\lambda_i' u\left(y_j;\theta\right)} \quad (2.32)$$

where

$$\sum_{j=1}^{n}\frac{w_{ij}u\left(y_j;\theta\right)}{1+\lambda_i' u\left(y_j;\theta\right)} = 0, \quad (2.33)$$

for each θ and $i, j = 1,\cdots,n$. Now we can form the local empirical log-likelihood (LELL) function at θ as

$$\mathbf{LELL}(\theta) = \sum_{i=1}^{n}\sum_{j=1}^{n} T_{i,n}w_{ij}\log\widehat{p}_{ij} = \sum_{i=1}^{n}\sum_{j=1}^{n} T_{i,n}w_{ij}\log\left\{\frac{w_{ij}}{1+\lambda_i' u\left(y_j;\theta\right)}\right\} \quad (2.34)$$

where λ_i solves (2.33) and $T_{i,n}$ is a sequence of trimming functions to deal with the so-called "denominator problem" (Robinson, 1987; Newey, 1993; **KTA**). Simply speaking, the positive weights we use, w_{ij}, are equal to $\mathcal{K}_{ij}/\left(nb_n^s\widehat{h}(x_i)\right)$ where $\widehat{h}(x_i) = \sum_{j=1}^{n}\mathcal{K}_{ij}/\left(nb_n^s\right)$ is a Nadaraya-Watson estimator of the marginal density $h(x_i)$ for the process $\{x_t\}$. When the data point x_i lies in the tails of $h(\cdot)$, w_{ij} and hence $\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}\log\widehat{p}_{ij}$ may be ill-behaved in the sense that no solution exists for the maximization of $\mathbf{LELL}(\theta)$ in (2.34). By trimming away small values of $\widehat{h}(x_i)$ via $T_{i,n}$, defined in this paper as $T_{i,n} = 1\left\{\widehat{h}(x_i) > b_n^{\varsigma}\right\}$ with $\varsigma \in (0,1)$ following **KTA**, the optimization in (2.34) will be well behaved.

Now, we can define the LEL estimator for the model (2.26) as:

$$\widehat{\theta}_{LEL} = \arg\max_{\theta\in\Theta} \mathbf{LELL}(\theta) \tag{2.35}$$

The implementation of the proposed estimator $\widehat{\theta}_{LEL}$ is straightforward by noticing

$$\lambda_i = \arg\max_{\gamma} \sum_{j=1}^{n} w_{ij} \log\left[1 + \gamma' u\left(y_j;\theta\right)\right]$$

from (2.5). Then the problem (2.34) can be equivalently stated as maximizing

$$\sum_{i=1}^{n}\sum_{j=1}^{n} T_{i,n} w_{ij} \log\left\{\frac{1}{1+\lambda_i' u\left(y_j;\theta\right)}\right\} = -\sum_{i=1}^{n} T_{in}\left\{\max_{\gamma}\sum_{j=1}^{n} T_{i,n} w_{ij}\log\left[1+\gamma' u\left(y_j;\theta\right)\right]\right\}$$

with respect to θ. Therefore, the numerical optimization can be performed with two loops by simple Newton-Raphson procedures: the "inner loop" with respect to γ and the "outer loop" with respect to θ.

Note that the construction of the LEL estimator for the model (2.26) assumes $u(\cdot)$ is explicitly defined, which is not the case for the Markov model identified by (2.11). To estimate our continuous-time Markov model (2.1), we replace $u(\cdot)$ by $\widetilde{u}(\cdot)$ defined by applying the approximation schemes proposed in Section 2.2.1. Correspondingly, we have the approximated local empirical log-likelihood (ALELL) function at θ as

$$\mathbf{ALELL}(\theta) = \sum_{i=1}^{n}\sum_{j=1}^{n} T_{i,n} w_{ij} \log\left\{\frac{w_{ij}}{1+\lambda_i' \widetilde{u}\left(y_j;\theta\right)}\right\}$$

Then, the LEL estimator for the Markov model (2.1) is defined as:

$$\widehat{\theta}_{ALEL} = \arg\max_{\theta\in\Theta} \mathbf{ALELL}(\theta) \tag{2.36}$$

2.2.3 Asymptotic Properties

In this section, I first assume $u(\cdot)$ is explicitly defined and derive asymptotic properties of the LEL estimator in (2.35) for the general case. Then the LEL estimator in (2.36) for the continuous-time Markov model (2.1) is treated. First we set up some notations used in the rest of the paper: C is a generic positive constant, $S^a = \{\xi \in \mathbb{R}^a : \|\xi\| = 1\}$ is the unit sphere in $\mathbb{R}^a$, $x^{(i)}$ denotes the *ith* component of the vector x and $M^{(i,j)}$ is the (i, j)th element of a matrix M. ∇_θ is the gradient operator with respect to θ; for example, $\nabla_\theta u(y;\theta) = \partial u'(y;\theta)/\partial\theta$, where $\partial u'(y;\theta)/\partial\theta$ is the transpose of the $q \times p$ matrix $\partial u(y;\theta)/\partial\theta$. If $f(\theta)$ is a scalar function, then $\nabla_\theta f(\theta)$ is a $p \times 1$ vector while the Hessian $\nabla_{\theta\theta} f(\theta)$ is a $p \times p$ matrix. The following regularity conditions are imposed (note that Δ is suppressed when there is no confusion).

Assumption 2.2.2. The process $\{y_t, x_t\}_{t=0,1,2,\cdots}$ is a strictly stationary and m-th order Markov process which is strongly mixing with mixing coefficient α_j satisfying

$$\alpha_j \le C j^{-\beta} \tag{2.37}$$

where for some $s > 2$,

$$E\|y_0\|^s < \infty \tag{2.38}$$

and

$$\beta > \frac{2s-2}{s-2} \tag{2.39}$$

Assumption 2.2.3. The marginal density of x_t, $h(x)$ satisfies

$$0 < h(x) \le \sup_{x\in\mathbb{R}^{dm}} h(x) \le C \tag{2.40}$$

and

$$\sup_{x \in \mathbb{R}^{dm}} E\left[\|u(y_0)\|^s |x_0 = x\right] h(x) \le C \tag{2.41}$$

Furthermore, there is some $j^* < \infty$ such that for all $j \ge j^*$

$$\sup_{x_0, x_j \in \mathbb{R}^{dm}} E\left[\left\|u(y_0)\, u\left(y_j\right)\right\| |x_0, x_j\right] h_j\left(x_0, x_j\right) \le C \tag{2.42}$$

where $h_j\left(x_0, x_j\right)$ denotes the joint density of $\left(x_0, x_j\right)$.

Assumption 2.2.4. For each $\theta \neq \theta_0$, there exists a set $\mathcal{X}_\theta \subset \mathbb{R}^s$ such that $P\{x \in \mathcal{X}_\theta\} > 0$ and $E\left[u(y;\theta)|x\right] \neq 0$ for every $x \in \mathcal{X}_\theta$.

Assumption 2.2.2 specifies that the serial dependence in the data is strongly mixing with exponentially decaying mixing coefficients satisfying (2.37)-(2.39), which are used to invoke the central limit theorem for U-statistics with weakly dependent data in Fan and Li (1999). Moreover, the m-th order Markov process, of which the model (2.1) is a special case in light of (2.11), enables us to make use of the semi-parametric efficiency bound in Carrasco and Florens (2008). In fact, Assumption 2.2.2 is a common regularity condition in econometric studies of continuous-time Markov models (Ait-Sahalia, 2002, 2008; Ait-Sahalia, Fan and Peng, 2009; Yu, 2007); see Ait-Sahalia (1996b), Ait-Sahalia and Mykland (2004, Lemma 4), Chen, Hansen, and Carrasco (2010), and Hansen and Scheinkman (1995) for primitive conditions and proofs.

Assumption 2.2.3 requires that the density $h(x)$ is bounded and (2.41)-(2.42) control the tail behaviors of the conditional expectations $E\left[\|u(y_0)\|^s |x_0 = x\right]$ and $E\left[\left\|u(y_0)\, u\left(y_j\right)\right\| |x_0, x_j\right]$ respectively. Assumption 2.2.3 prepares conditions for us to employ uniform convergence rates in Hansen (2008). Assumption 2.2.4 is the identification condition of θ_0 for the model (2.26). For our Markov model (2.1) as a special case of (2.26), Assumption 2.2.4 is implied by Assumption 2.2.1

Assumption 2.2.5. $E\left[\sup_{\theta}\|u(y;\theta)\|^{r}\right]<\infty$ for some $r\geq 8$.

The value r=8, following **KTA**, is used in the proof of Lemma A.5?.

Assumption 2.2.6. For $x=\left(x^{(1)},\cdots,x^{(dm)}\right)$, let $\mathcal{K}(x)=\Pi_{i=1}^{dm}\kappa\left(x^{(i)}\right)$. Here

(i), $\kappa(\cdot):\mathbb{R}\to\mathbb{R}$ is a continuously differentiable p.d.f. with support $[-1,1]$. $\kappa(\cdot)$ is symmetric about the origin, and for some $a\in(0,1)$ is bounded away from zero on $[-a,a]$.

(ii), For all $u,u'\in\mathbb{R}^{dm}$

$$\left|\mathcal{K}(u)-K(u')\right|\leq C\|u-u'\| \tag{2.43}$$

Assumption 2.2.6 imposes regularity conditions on the kernel function, allowing for most commonly used kernels, although the uniform and Dirichlet kernels are excluded. Since $\kappa(\cdot)$ is continuously differentiable with a bounded support, it is straightforward to show that $\mathcal{K}(u)$ is bounded and integrable, i.e.,

$$|\mathcal{K}(u)|\leq C and \int_{\mathbb{R}^{dm}}|\mathcal{K}(u)|\,du\leq C \tag{2.44}$$

Assumption 2.2.6 allows us to use the uniform convergence rates in Hansen (2008). The requirement that $\mathcal{K}$ is bounded away from zero on a closed ball centered at the origin is imposed in proving a modified version of Devroye and Wagner (1980, Lemma 2) for the proof of Lemma A.1.?

Assumption 2.2.7

(i), The marginal density of x_t, $h(x)\in C^2\left(\mathbb{R}^{dm}\right)$, $\sup_{x\in\mathbb{R}^{dm}}\|\nabla_x h(x)\|<\infty$, and $\sup_{x\in\mathbb{R}^{dm}}\|\nabla_{xx}h(x)\|<\infty$.

(ii), $u(y;\theta)$ is continuous in θ w.p.1 and $E\left\{\sup_{\theta\in\Theta}\|\nabla_\theta u(y;\theta)\|\right\}<\infty$.

(iii), $\left\|\nabla_{xx}\left\{h(x)E\left[u^{(l)}(y;\theta)|x\right]\right\}\right\|$ is uniformly bounded on $\Theta\times\mathbb{R}^{dm}$ for $1\le l\le q$.

Part (i) is used in the proof of Lemma A.2 and Proposition A.7 while parts (ii) and (iii) are useful for the consistency of $\widehat{\theta}_{LEL}$.

Assumption 2.2.8 There exists a closed ball B_0 around θ_0 such that for $1\le i,l\le q$ and $1\le j,k\le p$:

(i), $D(x;\theta)$ and $V(x;\theta)$ are continuous on B_0 w.p.1.

(ii), $\inf_{(\xi,x,\theta)\in S^q\times\mathbb{R}^{dm}\times B_0}\xi' V(x;\theta)\xi>0$ and $\sup_{(\xi,x,\theta)\in S^q\times\mathbb{R}^{dm}\times B_0}\xi' V(x;\theta)\xi<\infty$.

(iii), $\sup_{\theta\in B_0}\left\|\partial u^{(i)}(y;\theta)/\partial\theta^{(j)}\right\|\le d(y)$ and $\sup_{\theta\in B_0}\left\|\partial^2 u^{(i)}(y;\theta)/\left(\partial\theta^{(j)}\partial\theta^{(k)}\right)\right\|\le l(y)$ hold w.p.1 for some real-valued functions $d(y)$ and $l(y)$ such that $Ed^{\eta}(y_t)<\infty$ for $\eta\ge 4$ and $El(y_t)<\infty$.

(iv), $\sup_{x\in\mathbb{R}^{dm}}\left\|\nabla_x\left\{h(x)D^{(ij)}(x;\theta_0)\right\}\right\|<\infty$ and $\sup_{(x,\theta)\in\mathbb{R}^{dm}\times B_0}\left\|\nabla_{xx}\left\{h(x)D^{(ij)}(x;\theta_0)\right\}\right\|<\infty$.

(v), $\sup_{x\in\mathbb{R}^{dm}}\left\|\nabla_x\left\{h(x)V^{(il)}(x;\theta_0)\right\}\right\|<\infty$ and $\sup_{(x,\theta)\in\mathbb{R}^{dm}\times B_0}\left\|\nabla_{xx}\left\{h(x)V^{(il)}(x;\theta_0)\right\}\right\|<\infty$.

Assumption 2.2.8 is used in the proof of Proposition A.7 and Lemmas A.4-A.5. ?

Assumption 2.2.9. When solving (2.33) for $\lambda_i,\cdots,\lambda_n$, we only search over the set $\left\{\gamma\in\mathbb{R}^q:\|\gamma\|\le Cn^{-1/r}\right\}$ where r is as in Assumption A.4.

This is similar to Assumption 3.6 of **KTA** and only needed to establish the asymptotic normality of $\widehat{\theta}_{LEL}$. It is reasonable since the $\lambda_i's$ converge to zero under (2.26). Therefore, when solving (2.33) for $\lambda_i's$, we can search for the solution only in some neighborhood of the origin[11].

Assumption 2.2.10. $b_n \to 0$, $nb_n^{dm} \to \infty$, $n^{1/2}b_n^2/b_n^{\varsigma} \to 0$, $n^{1/r+1/2}\ln n/\left(nb_n^{dm+2\varsigma}\right) \to 0$, $n^{1/r+1/2}b_n^4/b_n^{2\varsigma} \to 0$, and $n\ln n/\left(nb_n^{dm+2\varsigma}\right) \to 0$.

Note that ς is from the trimming parameter $T_{i,n}$. $b_n \to 0$ and $nb_n^{dm} \to \infty$ are the standard conditions on the bandwidth to ensure the consistency of kernel estimators. The factor of $\ln n$ appears here since the uniform convergence rates for kernel estimators with dependent data in Hansen (2008) are employed.

We are now ready to present the consistency of $\widehat{\theta}_{LEL}$:

Theorem 2.2.1: Suppose Assumptions 2.2.1-2.2.8 and 2.2.10 hold and for some $\nu \geq dm$, the mixing coefficient β in Assumption 2.2.2 is restricted as

$$\beta > 1 + \frac{dm}{\nu} + dm, \tag{2.45}$$

and the bandwidth further satisfies

$$\frac{\ln n}{n^{\iota}b_n^{dm}} = o(1) \tag{2.46}$$

for

$$\iota = \frac{\beta - 1 - dm - \frac{dm}{\nu}}{\beta + 3 - dm}. \tag{2.47}$$

Assume further that

$$\sup_{x\in\mathbb{R}^{dm}} \|x\|^{\nu} h(x) \leq C, \tag{2.48}$$

$$\sup_{x\in\mathbb{R}^{dm}} \|x\|^{\nu} E\left[\|u(y_0)\|| x_0\right] h(x) \leq C \tag{2.49}$$

$$\sup_{x\in\mathbb{R}^{dm}} \|x\|^{\nu} E\left[\|\nabla_{\theta}u(y_0;\theta)\|| x_0\right] h(x) \leq C \tag{2.50}$$

[11] Similar to KTA, I did not impose such restrictions for both the simulation and empirical studies to check how tight these restrictions are in practice.

and

$$\sup_{x \in \mathbb{R}^{dm}} \|x\|^{\gamma} E\left[\left\| u\left(y_0;\theta\right) u\left(y_0;\theta\right)' \right\| \, x_0 \right] h\left(x\right) \leq C \tag{2.51}$$

Then as $n \to \infty$,

$$\widehat{\theta}_{LEL} \to^{a.s.} \theta_0$$

It can be seen from (2.45) and (2.47) that $\iota \in (0, 1]$ and thus (2.46) is a strengthening of the conventional requirement that $nb_n^{dm} \to \infty$. Conditions (2.48)-(2.51) are more restrictions on the tail behaviors of some conditional expectations used to invoke different uniform convergence rates in Hansen (2008).

Theorem 2.2.2: Suppose the same conditions under Theorem 2.2.1 and furthermore Assumptions 2.2.9 hold. Then as $n \to \infty$,

$$\sqrt{n}\left(\widehat{\theta}_{LEL} - \theta_0\right) \to^{d} N(0, I^{-1}\left(\theta_0\right))$$

$I^{-1}\left(\theta_0\right)$ coincides with the semi-parametric efficiency bound derived in Carrasco and Florens (2008) for m-th order Markov processes. Therefore, $\widehat{\theta}_{LEL}$ is asymptotically efficient. Such a result is new in the literature since existing results on the semi-parametric efficiency bound are only available for I.I.D data as in Newey (1990, 1993). The basic strategy of the proof for Theorems 2.2.1-2.2.2 is to modify the arguments in **KTA** to the current time-series context using uniform convergence rates in Hansen (2008) and a central limit theorem for U-statistics in Fan and Li (1999).

The asymptotic variance of $\widehat{\theta}_{LEL}$, i.e., $I^{-1}\left(\theta_0\right)$, has to be estimated consistently when conducting inferences for the individual parameters. Solving the maximization problem of $\mathcal{L}(\theta)$ with respect to $p_{i,j}, \mu_i$ and λ_j $(i, j = 1, \cdots, n)$ yields

$$\widehat{p}_{i,j}(\theta) = \frac{w_{ij} \exp\left[\widehat{\lambda}_i(\theta)' u\left(y_j;\theta\right)\right]}{\sum_{j=1}^{n} w_{ij} \exp\left[\widehat{\lambda}_i(\theta)' u\left(y_j;\theta\right)\right]} \tag{2.52}$$

which are estimates of the empirical conditional probabilities incorporating the restrictions from the model (2.26). Let $\widehat{p}_{i,j} = \widehat{p}_{i,j}\left(\widehat{\theta}_{LEL}\right)$, $\widehat{\lambda}_i = \widehat{\lambda}_i\left(\widehat{\theta}_{LEL}\right)$, $\widehat{u}_j = u\left(y_j;\widehat{\theta}_{LEL}\right)$, and $\widehat{U}_j = \partial u\left(y_j;\widehat{\theta}_{LEL}\right)/\partial\theta$. Then we can estimate $D(x_i)$ and $V(x_i)$ consistently by

$$\widehat{D} = \sum_{j=1}^{n} \widehat{p}_{i,j}\widehat{U}_j and \widehat{V} = \sum_{j=1}^{n} \widehat{p}_{i,j}\widehat{u}_j\widehat{u}_j' \tag{2.53}$$

respectively. The consistency of $\widehat{D}$ and $\widehat{V}$ can be shown by a straightforward extension of the proof in Smith (2007) for I.I.D data; I omit it here for brevity.

Now we deal with the impact of approximating numerical integrals in the estimator of (2.36) for our continuous-time Markov model (2.1)

Theorem 2.2.3: (i), Suppose the same conditions under Theorem 2.2.1 hold. Then

$$\widehat{\theta}_{ALEL} \rightarrow^{a.s.} \theta_0$$

as $n \to \infty$ and $M \to \infty$.

(ii), Suppose the same conditions under Theorem 2.2.2 hold. Then

$$\sqrt{n}\left(\widehat{\theta}_{ALEL} - \theta_0\right) \rightarrow^{d} N(0, I^{-1}(\theta_0))$$

as $n \to \infty$, $M \to \infty$, and $n^{1/2}/M \to 0$.

The condition $n^{1/2}/M \to 0$ reflects the interaction between the asymptotic schemes in discrete-time and continuous-time econometrics. Consistent asymptotic variance estimators of $\widehat{\theta}_{ALEL}$ can be obtained by replacing $u(\cdot)$ using $\widetilde{u}(\cdot)$ in (2.52)-(2.53).

2.3 Monte Carlo Simulations

In this section, I investigate the finite sample performances of the proposed LEL estimator for both univariate and multivariate continuous-time Markov processes. Specifically, I consider univariate diffusion and Levy jump models and multivariate diffusion and jump-diffusion models. For both univariate and multivariate diffusion models, comparisons will be made to MLE, either exact MLE (EMLE) or approximated MLE (AMLE) of Ait-Sahalia (2002, 2008), depending upon whether the likelihood function has a closed form. As can be seen from Section 2.2, choices of the trimming parameter ς, kernel function $\kappa(\cdot)$, and bandwidth b_n have to be made when computing the estimator $\widehat{\theta}_{LEL}$. First, similar to **KTA**, I set $T_{i,n} = 1$ for each i; i.e., I do not trim $\widehat{h}(\cdot)$[12]. Second, it is well known from the nonparametric estimation literature that choices of kernel functions do not change the results much. Henceforth, I use Bartlett kernel (Priestley, 1981) in all the simulations. Third, for the bandwidths b_n, which are usually the most important factors to be determined in nonparametric econometric inferences, I choose the cross-validation procedure suggested in Newey (1993). In fact, **KTA** find that in the I.I.D setting, the LEL estimator is relatively insensitive to the bandwidth choice and as such a cross-validation approach works well[13].

[12] I conduct simulation studies with different choices of the trimming parameter ς. Similar to **KTA**, I find that the performances of the LEL estimator for the current time series setting is not sensitive to ς.

[13] I also checked the performances of the LEL estimator with respect to different choices of bandwidths in the current time series setting. The results suggest that the performance is stable across a wide range of b_n's.

2.3.1 Diffusion Models

I first study the finite-sample performance of the proposed LEL estimator for both univariate and multivariate diffusion models. The following popular univariate diffusion models in modeling the short-rate dynamics are considered:

- DGP 2.1 (CIR Model):

$$dX_t = \kappa(\alpha - X_t)dt + \sigma\sqrt{X_t}dW_t,$$

  with $(\kappa, \alpha, \sigma^2) = (0.10, 0.08, 0.0004)$.

- DGP 2.2 (CKLS Model):

$$dX_t = \kappa(\alpha - X_t)dt + \sigma X_t^{\rho} dW_t,$$

  where $(\kappa, \alpha, \sigma, \rho) = (0.0972, 0.0808, 0.722399, 1.46)$.

- DGP 2.3 (Ait-Sahalia's (1996a) Nonlinear Drift Model):

$$dX_t = (\kappa_{-1}X_t^{-1} + \kappa_0 + \kappa_1 X_t + \kappa_2 X_t^2)dt + \sigma X_t^{\rho} dW_t,$$

  where $(\kappa_{-1}, \kappa_0, \kappa_1, \kappa_2, \sigma^2, \rho)$=(0.00107,-0.0517, 0.877, -4.604, 0.64754, 1.50).

DGP 2.1 has a closed-form scaled-χ^2 density. DGPs 2.2-2.3 both belong to the category of constant elastic variance (CEV) diffusion models and do not admit analytic transition densities. Therefore, DGP 2.1 can be estimated by EMLE while DGPs 2.2-2.3 by the AMLE in Ait-Sahalia (2002). Note also that the first two models have a linear drift but the third is nonlinear in the specification of the drift function. The parameter values I choose for all the univariate diffusion models are practically reasonable (see Ait-Sahalia (1996a), Hong and Li (2005), and Pritsker (1998)). For the CIR model, I specifically choose a small value for the mean-reverting parameter to check the performances of both LEL and MLE under a highly persistent data-generating process.

For multivariate diffusion models, I consider the following two-factor affine diffusions (Dai and Singleton, 2000; Ait-Sahalia and Kimmel, 2010):

- DGP 2.4: $A_0(2)$(or Bivariate O-U process):

$$d\begin{bmatrix} X_{1t} \\ X_{2t} \end{bmatrix} = \begin{bmatrix} \kappa_{11} & 0 \\ \kappa_{21} & \kappa_{22} \end{bmatrix} \begin{bmatrix} X_{1t} \\ X_{2t} \end{bmatrix} dt + d\begin{bmatrix} W_{1t} \\ W_{2t} \end{bmatrix}$$

  with W_{1t} and W_{2t} two independent Brownian Motions and $(\kappa_{11}, \kappa_{21}, \kappa_{22})$ =(-0.1117,1.1138,-1.1637).

- DGP 2.5: $A_1(2)$

$$d\begin{bmatrix} X_{1t} \\ X_{2t} \end{bmatrix} = \left(\begin{bmatrix} \alpha_1 \\ 0 \end{bmatrix} + \begin{bmatrix} \kappa_{11} & 0 \\ \kappa_{21} & \kappa_{22} \end{bmatrix} \begin{bmatrix} X_{1t} \\ X_{2t} \end{bmatrix} \right) dt + \begin{pmatrix} \sqrt{X_{1t}} & 0 \\ 0 & \sqrt{1 + \sigma_{21} X_{1t}} \end{pmatrix} d\begin{bmatrix} W_{1t} \\ W_{2t} \end{bmatrix}$$

  with W_{1t} and W_{2t} two independent Brownian Motions and $(\kappa_{11}, \kappa_{22}, \alpha_1, \kappa_{21}, \sigma_{21})$=(-0.7, -2.5, 0.56, 0.6, 0.5)

- DGP 2.6: $A_2(2)$(or Bivariate Feller)

$$d\begin{bmatrix} X_{1t} \\ X_{2t} \end{bmatrix} = \left(\begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} + \begin{bmatrix} \kappa_{11} & \kappa_{12} \\ \kappa_{21} & \kappa_{22} \end{bmatrix} \begin{bmatrix} X_{1t} \\ X_{2t} \end{bmatrix} \right) dt + \begin{pmatrix} \sqrt{X_{1t}} & 0 \\ 0 & \sqrt{X_{2t}} \end{pmatrix} d\begin{bmatrix} W_{1t} \\ W_{2t} \end{bmatrix}$$

  with W_{1t} and W_{2t} two independent Brownian Motions and (κ_{11}, κ_{12}, κ_{21}, κ_{22}, α_1, α_2) =(-0.7, 0.3, 0.4, -0.8, 0.56, 0.64).

DGP-2.4, the $A_0(2)$ model, is actually a special case of bivariate O-U processes and has a Gaussian transition density (Duffee, 2002) while neither DGP-2.5 nor DGP-2.6 has a closed-form transition density. Hence, DGP-2.4 can be estimated by EMLE and DGPs 5-6 by AMLE in Ait-Sahalia (2008). The parameter values for DGPs 2.4-2.6 are chosen either from Ait-Sahalia and Kimmel's (2010) empirical estimates using constructed yield data for US Treasury bonds from January of 1972 to December of 2002 or specifically to make the generated series positive when a square root term appears in diffusion functions, as in DGPs 2.5-2.6.

For each model, I simulate 1000 (the number of replications) data sets of a random sample $\{X_{t=\tau\Delta}\}_{\tau=1}^{n}$ at the monthly frequency ($\Delta = 1/12$) for n=250, 500, and 1000 respectively. These are obtained by first simulating daily data from the models and then recording monthly data with 22 daily observations in each month. That is, we have $M = 22$ in terms of the asymptotic schemes in Section 3.1 for the approximation of the numerical integrals. These sample sizes correspond to up to around 100 years of monthly data. For each replication, I estimate the model using LEL and MLE. To evaluate the impact of the approximation of the numerical integrals relative to the direct discretization of the models, estimators based on the Euler schemes are also computed. For example, the Euler discretization of DGP 2.1 is defined as

$$X_{t+\Delta} - X_t = \kappa(\alpha - X_t)\Delta + \sigma\sqrt{X_t}\sqrt{\Delta}\varepsilon_{t+\Delta},$$

where $\varepsilon_{t+\Delta} \sim N(0,1)$ so that the transition density for this discretized model is

$$\begin{aligned} & p\left(X_t = x | X_{t-\Delta} = x_0\right) \\ = \; & \left(2\pi\Delta\sigma^2 X_t\right)^{-1/2} \exp\left\{-\left(x - x_0 - \kappa(\alpha - X_t)\Delta\right)^2 / \left(2\Delta\sigma^2 X_t\right)\right\} \end{aligned}$$

For the resulting series of estimators, the empirical bias and root mean squared error (RMSE) are reported. The data generation for each model is carried out by the following steps. First, an initial value X_0 is drawn from the marginal density if it is available in closed form or set at at the long-run mean of the model. Second, given a value X_t, we generate $X_{t+\Delta}$ according to the transition density if it is available or the Euler-Milstein scheme if not. To mitigate the impact of initial values, 500 more data points are generated as the "burn-in" period.

The results for univariate DGPs 2.1-2.3 are reported in Table 2.1 with both LEL and MLE reported. In particular, results of Euler estimators are reported

Table 2.1: **Comparison of LEL with MLE and Euler Schemes for Univariate Diffusion Models**

Parameter	n=250		n=500		n=1000	
Estimate	Bias	RMSE	Bias	RMSE	Bias	RMSE
	DGP 2.1					
$\widehat{\kappa}_{LEL}$	0.1023	0.1746	0.0432	0.0922	0.0210	0.0648
$\widehat{\alpha}_{LEL}$	0.0018	0.0300	0.0017	0.0083	0.0012	0.0059
$\widehat{\sigma}^2_{LEL}$	0.0002	$4.30{\cdot}10^{-4}$	0.0001	$2.37{\cdot}10^{-4}$	$7.26{\cdot}10^{-5}$	$1.68{\cdot}10^{-4}$
$\widehat{\kappa}_{EMLE}$	0.0812	0.1136	0.0409	0.0831	0.0343	0.0624
$\widehat{\alpha}_{EMLE}$	0.0036	0.0060	0.0039	0.0065	0.0028	0.0052
$\widehat{\sigma}^2_{EMLE}$	$1.38{\cdot}10^{-6}$	$2.55{\cdot}10^{-5}$	$5.73{\cdot}10^{-5}$	$7.98{\cdot}10^{-5}$	$7.13{\cdot}10^{-7}$	$1.48{\cdot}10^{-5}$
$\widehat{\kappa}_{Euler}$	0.1806	0.3033	0.1066	0.1442	0.0635	0.1193
$\widehat{\alpha}_{Euler}$	0.0154	0.0587	0.0082	0.0277	0.0067	0.0104
$\widehat{\sigma}^2_{Euler}$	0.0004	$9.11{\cdot}10^{-4}$	0.0002	$8.56{\cdot}10^{-4}$	$1.23{\cdot}10^{-4}$	$5.31{\cdot}10^{-4}$

	DGP 2.2					
$\widehat{\kappa}_{LEL}$	0.0155	0.0447	0.0143	0.0379	0.0176	0.0269
$\widehat{\alpha}_{LEL}$	0.0475	0.1158	0.0232	0.0774	0.0111	0.0640
$\widehat{\sigma}_{LEL}$	-0.2065	0.4488	-0.1540	0.4059	-0.0998	0.4014
$\widehat{\rho}_{LEL}$	-0.2394	0.6793	-0.1995	0.5505	-0.1554	0.4932
$\widehat{\kappa}_{AMLE}$	0.0890	0.1224	0.0711	0.0984	0.0594	0.0854
$\widehat{\alpha}_{AMLE}$	0.0941	1.0651	0.0332	0.4296	0.0290	0.3184
$\widehat{\sigma}_{AMLE}$	0.1180	0.4417	0.0386	0.2193	0.0154	0.1545
$\widehat{\rho}_{AMLE}$	0.0063	0.1086	0.0022	0.0648	-0.0014	0.0519
	DGP 2.3					
$\widehat{\kappa}_{-1,LEL}$	-0.0004	$8.40 \cdot 10^{-4}$	-0.0003	$8.37 \cdot 10^{-4}$	-0.0003	$6.86 \cdot 10^{-4}$
$\widehat{\kappa}_{0,LEL}$	-0.0097	0.0374	-0.0009	0.0360	-0.0008	0.0412
$\widehat{\kappa}_{1,LEL}$	-0.2181	0.4884	-0.2190	0.4502	-0.2039	0.4644
$\widehat{\kappa}_{2,LEL}$	0.0751	3.3601	0.0581	3.0237	0.0938	2.5865
$\widehat{\sigma}^2_{LEL}$	0.1230	0.5763	0.0514	0.4077	0.0356	0.7644
$\widehat{\rho}_{LEL}$	-0.4440	1.1851	-0.4348	1.1786	-0.4189	1.3123
$\widehat{\kappa}_{-1,AMLE}$	0.0006	0.0033	0.0004	0.0016	0.0003	0.0012
$\widehat{\kappa}_{0,AMLE}$	0.0078	0.0871	0.0024	0.0700	-0.0083	0.0624
$\widehat{\kappa}_{1,AMLE}$	-0.1331	1.0870	-0.0563	0.9554	0.1169	0.8953
$\widehat{\kappa}_{2,AMLE}$	-1.0268	7.2619	-0.7932	5.6634	-0.9381	4.3739
$\widehat{\sigma}^2_{AMLE}$	0.0377	0.5651	0.0137	0.3697	0.0541	0.2851
$\widehat{\rho}_{AMLE}$	-0.0875	0.2240	-0.0682	0.1913	-0.0166	0.1341

Notes: (i), The modes are DGP 2.1, CIR model $dX_t = \kappa(\alpha - X_t)dt + \sigma\sqrt{X_t}dW_t$, DGP 2.2, CKLS model $dX_t = \kappa(\alpha - X_t)dt + \sigma X_t^{\rho} dW_t$, and DGP 2.3, Ait-Sahalia's Nonlinear Drift Model, $dX_t = (\kappa_{-1}X_t^{-1} + \kappa_0 + \kappa_1 X_t + \kappa_2 X_t^2)dt + \sigma X_t^{\rho} dW_t$. (ii), The sampling frequency is monthly with $\Delta = 1/22$. (iii), The number of replications is 1000.

for DGP 2.1[14]. For the CIR model, LEL is as good as EMLE for the mean reverting speed κ and long-run mean α in terms of both the empirical bias and RMSE[15]. The latter actually confirms the efficiency of the proposed LEL. For the volatility coefficient, EMLE has superior performance, but the performance of LEL improves very quickly as the sample size increases. In addition, both LEL and EMLE outperform the Euler estimators. For the CKLS model, LEL seems to perform better than AMLE for both κ and α for any sample size. For example, when the sample size n=500, the bias of EMLE for κ is 5 times bigger than that of LEL. Of course, EMLE has superior performance for the volatility parameters σ and ρ. For Ait-Sahalia's general nonlinear drift model, LEL performs a bit better than AMLE for all the drift parameters except κ_1. For example, for the coefficient of the squared term κ_2, LEL has a bias of around 0.09 while that of AMLE is is about -0.9 when n is 1000. Similar to DGPs 2.1-2.2, AMLE performs better than LEL for diffusion parameters.

[14] I compute the Euler estimators for all diffusion models considered in the simulation studies. Since results are very similar, I only report the performances of Euler estimators for DGP 1 to save space.

[15] Observe that both LEL and EMLE have large finite-sample biases under the highly persistent CIR model in DGP1; the percentage bias is almost 100% for both estimators when n is 250. In fact, this severely poor finite-sample performance of mean reverting speed estimator has been discovered as early as Merton (1980) and studied by Phillips and Yu (2005) and Tang and Chen (2009). The main reason is that for diffusion models, the drift term can only be identified when the time span $T \to \infty$ (see Bandi and Phillips (2003) for details). I conjecture that methods to reduce the finite sample bias can be combined with LEL more conveniently thant MLE due to the closed-form property. But I do not explore this direction here since it is out of the current focus.

The results for multivariate DGPs-2.4-2.6 are reported in Table 2.2. For the $A_0(2)$ model, LEL performs similar to EMLE, except that the former is improving faster than the latter when n is increased. For example, both LEL and EMLE have biases around -0.04 for the mean reverting parameter κ_{11} of the first component process when the n is equal to 250. As the sample size is increased to 1000, the EMLE's bias is about 0.03 while that of LEL is only -0.015. Another observation is that for the parameter κ_{21} controlling the correlation between X_{1t} and X_{2t}, LEL is performing appreciably better than EMLE. For both $A_1(2)$ and $A_2(2)$ models, the patterns of finite-sample performance are similar to DGPs 2.2-2.3 in that LEL performs better than the AMLE for drift parameters while the scenario is reversed for diffusion parameters. Finally, for the parameter controlling the correlation between the two components in the $A_1(2)$ model, LEL performs better than AMLE. However, for the $A_2(2)$ model with two parameters κ_{12} and κ_{21} capturing the correlation, LEL is better for the first while AMLE is better for the second.

Table 2.2: **Comparison of LEL with MLE for Multivariate Diffusion Models**

Parameter	$n = 250$		$n = 500$		$n = 1000$	
Estimate	Bias	RMSE	Bias	RMSE	Bias	RMSE
	DGP 2.4: $(\kappa_{11}, \kappa_{21}, \kappa_{22})$ =(-0.1117,1.1138,-1.1637).					
$\widehat{\kappa}_{11,LEL}$	-0.0443	0.2005	-0.0335	0.1214	-0.0156	0.0728
$\widehat{\kappa}_{21,LEL}$	-0.0485	0.3656	-0.0540	0.2588	-0.0420	0.1826
$\widehat{\kappa}_{22,LEL}$	0.0238	0.3689	0.0328	0.2615	-0.0331	0.1820
$\widehat{\kappa}_{11,EMLE}$	-0.0412	0.3866	-0.0394	0.3166	0.0314	0.2558
$\widehat{\kappa}_{21,EMLE}$	-0.1066	0.0337	-0.1067	0.0317	-0.1066	0.0247

$\widehat{\kappa}_{22,EMLE}$	-0.0544	0.0268	-0.0648	0.0309	-0.0550	0.0258
	DGP 2.5: $(\kappa_{11}, \kappa_{22}, \alpha_1, \kappa_{21}, \sigma_{21})$ =(-0.7, -2.5, -0.56, 0.6, 0.5)					
$\widehat{\kappa}_{11,LEL}$	-0.2131	0.5370	-0.1171	0.3433	-0.1041	0.2488
$\widehat{\kappa}_{22,LEL}$	-0.3786	3.8944	-0.2035	1.1723	-0.0940	0.5030
$\widehat{\alpha}_{1,LEL}$	0.2233	0.4044	0.1517	0.2652	0.1415	0.2035
$\widehat{\kappa}_{21,LEL}$	0.1417	1.6476	0.0961	0.5426	0.0705	0.2998
$\widehat{\sigma}_{21,LEL}$	0.0485	2.0665	0.0364	0.5270	0.0170	0.2366
$\widehat{\kappa}_{11,EMLE}$	-0.4241	0.6044	-0.2072	0.4199	-0.1525	0.3499
$\widehat{\kappa}_{22,EMLE}$	-0.4000	2.2560	-0.1948	1.1313	-0.1336	0.7364
$\widehat{\alpha}_{1,EMLE}$	0.1455	0.3814	0.1072	0.2900	0.0902	0.2113
$\widehat{\kappa}_{21,EMLE}$	0.2821	1.9033	0.1332	0.6905	0.0684	0.4207
$\widehat{\sigma}_{21,EMLE}$	0.0100	0.8755	0.0054	0.1095	0.0021	0.0817
	DGP 2.6: $(\kappa_{11}, \kappa_{12}, \kappa_{21}, \kappa_{22}, \alpha_1, \alpha_2)$=(-0.7, 0.3, 0.4, -0.8, -0.56, -0.64)					
$\widehat{\kappa}_{11,LEL}$	-0.1141	0.47619	-0.0752	0.3188	-0.0340	0.2098
$\widehat{\kappa}_{12,LEL}$	-0.0505	0.45339	-0.0050	0.2836	-0.0070	0.1795
$\widehat{\kappa}_{21,LEL}$	0.0209	0.48656	0.0037	0.3137	-0.0053	0.1899
$\widehat{\kappa}_{22,LEL}$	-0.2376	0.55290	-0.0929	0.3283	-0.0332	0.2092
$\widehat{\alpha}_{1,LEL}$	0.2325	0.67476	0.1306	0.4213	0.0677	0.2832
$\widehat{\alpha}_{2,LEL}$	0.3572	0.77610	0.1683	0.4796	0.0758	0.2892
$\widehat{\kappa}_{11,EMLE}$	-0.5662	1.10367	-0.3416	0.6991	-0.1442	0.4812
$\widehat{\kappa}_{12,EMLE}$	-0.1252	0.61860	-0.0771	0.5123	-0.0265	0.1160
$\widehat{\kappa}_{21,EMLE}$	0.0110	0.38030	0.0013	0.1506	-0.0009	0.0541
$\widehat{\kappa}_{22,EMLE}$	-0.6555	0.84404	-0.3100	0.5326	-0.1083	0.3502
$\widehat{\alpha}_{1,EMLE}$	0.1677	0.58652	0.0625	0.3159	0.0254	0.1927
$\widehat{\alpha}_{2,EMLE}$	0.2613	0.63269	0.1148	0.3912	0.0221	0.1883

Notes: (i), The modes are DGP 2.4, $A_0(2)$ model, DGP 2.5, the correlated $A_1(2)$ and DGP 2.6, $A_2(2)$ model respectively (ii), The sampling frequency is monthly with $\Delta = 1/22$. (iii), The number of replications is 1000.

2.3.2 Jump-Diffusion Models

I consider the following processes:

- DGP 2.7: CKLS-P model in (5.7) with $(\kappa, \alpha, \sigma, \rho, \lambda, \mu_y, \sigma_y)$=(1.0252, $1.1\cdot 10^{-4}$, 0.0088, 1.2110, 0.0106, -0.0210, 0.0068)
- DGP 2.8: Bivariate CIR Jump-Diffusion model

$$d\begin{bmatrix} X_{1t} \\ X_{2t} \end{bmatrix} = \left(\begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} + \begin{bmatrix} \kappa_{11} & \kappa_{12} \\ \kappa_{21} & \kappa_{22} \end{bmatrix}\begin{bmatrix} X_{1t} \\ X_{2t} \end{bmatrix}\right)dt + \begin{pmatrix} \sigma_{11}\sqrt{X_{1t}} & 0 \\ 0 & \sigma_{22}\sqrt{X_{2t}} \end{pmatrix} d\begin{bmatrix} W_{1t} \\ W_{2t} \end{bmatrix} + JdN_t,$$

  where W_{1t} and W_{2t} two independent Brownian Motions, (κ_{11}, κ_{12}, κ_{21}, κ_{22}, α_1, α_2, σ_{11}^2, σ_{22}^2)=(-0.7, 0.3, 1.2, -0.8, -0.56, -0.48, 0.002, 0.001), J is the random jump size which follows a $N(\mu_J, \Omega_J)$ distribution with $\mu_J = (0.05, 0.01)$ and $\Omega_J = \begin{pmatrix} 0.2 & 0 \\ 0 & 0.12 \end{pmatrix}$, and N_t is a Poisson process with arrival intensity $\lambda = 5$.

DGP-2.7 is the CKLS model in DGP-2.2 augmented by a jump term. The parameter values are chosen from the empirical estimates in Section 2.4 using daily Euro/Dollar rates from June 25, 2000 to June 25 2010 to be practically reasonable. DGP-2.8 is a Bivariate CIR model augmented by a jump term JdN_t. The specification of the jump intensity implies that both X_{1t} and X_{2t} jump together, i.e., the jump times are the same for the two components[16]. For the jump size distribution, we can see that the jump sizes for the two component processes are different and furthermore independent of each other. Correlated jumps can be introduced by assigning non-zero values for the cross-terms in Ω_J. The data-generating schemes for DGPs 2.7-2.8 are similar to those for DGPs 2.1-2.6, with 1000 replications, daily frequency ($\Delta = 1/252$, four out of the five data points sampled a day are discarded, and the one left is recorded as the daily observation) for n=1000, 2500, and 5000 corresponding to up to 20 years, Euler-Milstein discretization, and 500 more data points as the "burn-in" period. Note M is set to 1 here, and the simulation results can show the impact of the numerical approximation errors for daily data.

To evaluate whether the choice of $f(\cdot)$ as the exponential function in (2.13) results in large loss of identification information of model dynamics, I also compute the simulated MLE (SMLE) in Brandt and Santa-Clara (2002) and Piazzesi (2005) for comparison.[17] The SMLE proposed in Brandt and Santa-Clara (2002) is only for diffusion models, and Piazzesi (2005) extends it to jump-diffusion models. The simulated transition density is obtained as follows. First, the time interval Δ between any two consecutive observations are split into smaller intervals (Δ/q) and a high-frequency discrete-time process is simulated based on the Euler discretization scheme. Second, the simulation in the first step is repeated S times, and transition densities between two consecutive observations are computed by the average of the S simulated paths. Following Brandt and Santa-Clara (2002), Durham and Gallant (2001), and Piazzesi (2005), I set $q = 10$

[16]It can be seen from the identification condition (2.10) that the proposed LEL estimator can be extended straightforwardly to cases with asynchronous jumps.

[17]The SMLE approach is orginally developed by Stata-Clara (1995) in an earlier version of Brandt and Santa-Clara (2002) and by Pedersen (1995) independently.

Table 2.3: **Comparison of LEL with MLE for Univariate Jump Diffusion Models**

Parameter	$n = 1000$		$n = 2500$		$n = 5000$	
Estimate	Bias	RMSE	Bias	RMSE	Bias	RMSE
	DGP 2.7					
$\widehat{\kappa}_{LEL}$	0.5226	1.1062	0.3874	0.4038	0.1655	0.2255
$\widehat{\alpha}_{LEL}$	$2.28{\cdot}10^{-5}$	$1.2{\cdot}10^{-4}$	$1.46{\cdot}10^{-5}$	$4.3{\cdot}10^{-5}$	$1.12{\cdot}10^{-5}$	$2.6{\cdot}10^{-5}$
$\widehat{\sigma}_{LEL}$	0.0020	0.0071	0.0013	0.0053	0.0002	0.0022
$\widehat{\rho}_{LEL}$	0.3990	0.8890	0.2025	0.3979	0.0922	0.2806
$\widehat{\lambda}_{LEL}$	0.0024	0.0168	0.0012	0.0134	0.0008	0.0056
$\widehat{\mu}_{y,LEL}$	0.0047	0.0091	0.0034	0.0075	0.0010	0.0044
$\widehat{\sigma}_{y,LEL}$	0.0015	0.0070	0.0011	0.0037	0.0004	0.0019
$\widehat{\kappa}_{SMLE}$	0.5113	0.9964	0.3523	0.3888	0.1642	0.2188
$\widehat{\alpha}_{SMLE}$	$2.02{\cdot}10^{-5}$	$9.4{\cdot}10^{-5}$	$1.41{\cdot}10^{-5}$	$2.6{\cdot}10^{-5}$	$1.07{\cdot}10^{-5}$	$1.1{\cdot}10^{-5}$
$\widehat{\sigma}_{SMLE}$	0.0017	0.0066	0.0009	0.0041	0.0002	0.0015
$\widehat{\rho}_{SMLE}$	0.3781	0.8903	0.1767	0.3832	0.0806	0.2667
$\widehat{\lambda}_{SMLE}$	0.0026	0.0157	0.0011	0.0128	0.0006	0.0051
$\widehat{\mu}_{y,SMLE}$	0.0051	0.0085	0.0032	0.0070	0.0011	0.0038
$\widehat{\sigma}_{y,SMLE}$	0.0022	0.0076	0.0013	0.0041	0.0005	0.0018

Notes: (i), The modes is DGP-2.7, the CKLS-P model. (ii), The sampling frequency is daily with $\Delta = 1/252$. (iii), The number of replications is 1000.

and $S = 5000$ in the simulation studies.

The results for LEL and SMLE of DGPs 2.7-2.8 are reported in Table 2.3-2.4. It can be seen that LEL has comparable performances to SMLE for both the CKLS-P and bivariate CIR Jump-Diffusion models, especially for the jump parameters.

For CKLS-P, the percentage biases for the jump intensity λ, jump size mean and standard deviation are 22.64%, 22.38% and 22.06% respectively for n=1000 and reduced to 7.55%, 4.76% and 5.88% respectively when n is equal to 5000. Similar performances can be observed for the jump parameters of DGP 2.8. Last, large biases similar to those for DGPs 2.1-2.3 still exist for the mean reverting parameter. Overall, simulation evidence reveals that numerical approximation errors are negligible even when $M = 1$ in finite samples. Moreover, the observed comparable performances of LEL to SMLE in terms of the RMSE of parameter estimates suggests that the choice of $f(\cdot)$ as the exponential function in (2.13) does not lead to large identification information loss of model dynamics.

Table 2.4: **Comparison of LEL with MLE for Multivariate Jump Diffusion Models**

Parameter Estimate	$n = 1000$ Bias	$n = 1000$ RMSE	$n = 2500$ Bias	$n = 2500$ RMSE	$n = 5000$ Bias	$n = 5000$ RMSE
	DGP 2.8					
$\widehat{\kappa}_{11,LEL}$	-0.3140	0.4692	-0.2263	0.1999	-0.1005	0.0875
$\widehat{\kappa}_{12,LEL}$	-0.0972	0.3006	-0.0501	0.1122	-0.0344	0.0776
$\widehat{\kappa}_{21,LEL}$	0.4988	1.0073	0.2836	0.0889	0.1220	0.0702
$\widehat{\kappa}_{22,LEL}$	-0.3504	0.6068	-0.1429	0.4225	-0.0667	0.1991
$\widehat{\alpha}_{1,LEL}$	0.0773	0.4842	0.0405	0.2754	0.0235	0.1661
$\widehat{\alpha}_{2,LEL}$	0.0550	0.5005	0.0371	0.2226	0.0226	0.1790
$\widehat{\sigma}^2_{11,LEL}$	0.0004	0.0018	0.0002	0.0014	0.0001	0.0011
$\widehat{\sigma}^2_{22,LEL}$	0.0002	0.0013	$8.5{\cdot}10^{-5}$	0.0008	$3.3{\cdot}10^{-5}$	0.0004
$\widehat{\lambda}_{,LEL}$	0.5269	4.8874	0.3744	3.0512	0.1333	2.1996
$\widehat{\mu}_{J1,LEL}$	0.0118	0.0503	0.0068	0.0333	0.0042	0.0177

$\widehat{\mu}_{J2,LEL}$	0.0027	0.0086	0.0015	0.0065	0.0008	0.0051
$\widehat{\Omega}_{J11,LEL}$	0.0366	0.2314	0.0157	0.1667	0.0061	0.0988
$\widehat{\Omega}_{J22,LEL}$	0.0286	0.1402	0.0131	0.0853	0.0055	0.0574
$\widehat{\kappa}_{11,SMLE}$	-0.3166	0.3885	-0.2002	0.1776	-0.0992	0.0833
$\widehat{\kappa}_{12,SMLE}$	-0.0935	0.2714	-0.0433	0.0982	-0.0277	0.0711
$\widehat{\kappa}_{21,SMLE}$	0.4253	0.8739	0.2227	0.0944	0.1066	0.0663
$\widehat{\kappa}_{22,SMLE}$	-0.3455	0.5777	-0.1228	0.3886	-0.0533	0.1741
$\widehat{\alpha}_{1,SMLE}$	0.0802	0.4736	0.0389	0.2722	0.0234	0.1618
$\widehat{\alpha}_{2,SMLE}$	0.0563	0.4352	0.0378	0.2139	0.0233	0.1795
$\widehat{\sigma}^2_{11,SMLE}$	0.0003	0.0016	0.0002	0.0009	0.0001	0.0007
$\widehat{\sigma}^2_{22,SMLE}$	0.0001	0.0011	$2.6 \cdot 10^{-5}$	0.0006	$8.2 \cdot 10^{-6}$	0.0002
$\widehat{\lambda}_{,SMLE}$	0.5333	4.6667	0.3756	3.0602	0.1305	2.2111
$\widehat{\mu}_{J1,SMLE}$	0.0120	0.0447	0.0064	0.0339	0.0039	0.0182
$\widehat{\mu}_{J2,SMLE}$	0.0025	0.0084	0.0013	0.0057	0.0007	0.0055
$\widehat{\Omega}_{J11,SMLE}$	0.0302	0.1995	0.0147	0.1554	0.0062	0.1006
$\widehat{\Omega}_{J22,SMLE}$	0.0244	0.1039	0.0127	0.0771	0.0051	0.0582

Notes: (i), The mode is DGP 2.8, the Bivariate CIR Jump Diffusion model which is the Bivariate CIR model in DGP 2.6 augmented by a jump term JdN_t where J is the random jump size which follows a $N(\mu_J, \Omega_J)$ distribution with $\mu_J = (\mu_{J1}, \mu_{J2})$ and $\Omega_J = \begin{pmatrix} \Omega_{J11} & 0 \\ 0 & \Omega_{J22} \end{pmatrix}$, N_t is a compound Poisson process with arrival intensity λ, and W_{1t} and W_{2t} are two independent Brownian motions. (ii), The sampling frequency is daily with $\Delta = 1/252$. (iii), The number of replications is 1000.

2.3.3 Levy Driven Jump-Diffusion Models

To check the finite sample performances of my LEL estimators for Levy-driven jump-diffusion models, I consider DGP 2.9, the CKLS-VG model in (2.58) with $(\kappa, \overline{\alpha}, \sigma^2, \rho, \gamma, \sigma_{VG}^2, \nu)$=(0.8426, 1.5·10^{-4}, 0.0024, 1.5300, -0.0182, 0.0069, 2.4900) and DGP 2.10, the CKLS-LS model in (2.59) with $(\kappa, \overline{\alpha}, \sigma^2, \rho, \sigma_{LS}, \alpha)$=(0.8223, 1.2·10^{-4}, 0.0522, 2.1400, 0.0308, 1.7115). These parameter values are also chosen from the empirical estimates in Section 2.4. The data-generating schemes for DGPs 2.9-2.10 are the same as those for DGPs 2.7-2.8, where the Euler discretizations of the models are

$$X_{t+\Delta} = X_t + \kappa\left(\overline{\alpha} - X_t\right)\Delta + \sigma X_t^{\rho}\varepsilon_{t+\Delta}^{VG}\sqrt{\Delta} + J_{t+\Delta}^{VG}$$

for CKLS-VG, where $\{\varepsilon_t^{VG}\}$ is $i.i.d.N(0,1)$, $J_{t+\Delta}^{VG} = \gamma G_{t+\Delta} + \sigma_{VG}\sqrt{G_{t+\Delta}}\nu_{t+\Delta}^{VG}$, $\nu_{t+\Delta}^{VG}$ is $i.i.d.N(0,1)$, $G_{t+\Delta}$ is $i.i.d.\Gamma\left(\frac{\Delta}{\nu}, \nu\right)$, and all the innovations are independent;

$$X_{t+\Delta} = X_t + \kappa\left(\overline{\alpha} - r_t\right)\Delta + \sigma X_t^{\rho}\varepsilon_{t+\Delta}^{LS}\sqrt{\Delta} + J_{t+\Delta}^{LS}$$

for CKLS-LS, where $\{\varepsilon_t^{LS}\}$ is $i.i.d.N(0,1)$, and $J_{t+\Delta}^{LS}$ is independent of $\{\varepsilon_t^{LS}\}$ and follows $S_\alpha\left(-1, \sigma_{LS}\Delta^{1/\alpha}, 0\right)$, a stable distribution with shape parameter α, skewness parameter -1, zero drift, and scale parameter $\sigma_{LS}\Delta^{1/\alpha}$.

Although the simulation of VG is straightforward, it is a little difficult to simulate data from LS because there are no standard random number generators for stable distributions. Following Li, Wells and Yu (2008), I employ the method of Chambers, Mallows and Stuck (1976) to simulate stable random variables through a nonlinear transformation of two independent uniform random variables. This method works for the arbitrary characteristic exponent $\alpha \in (0, 2)$ and the skewness parameter $\beta \in [-1, 1]$. When applying this method to the LS process, we set the skewness parameter β = -1 and transform the simulated

Table 2.5: **Simulation Results for Levy Driven Jump Diffusion Models**

	n=1000						n=2500				n=5000			
	True Values		CKLS-VG		CKLS-LS		CKLS-VG		CKLS-LS		CKLS-VG		CKLS-LS	
	CKLS-VG	CKLS-LS	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
κ	0.8426	0.8223	0.2026	0.6502	0.2104	0.6711	0.1145	0.3303	0.1160	0.3413	0.0508	0.1442	0.0515	0.1460
$\overline{\alpha}$	$1.5{\cdot}10^{-4}$	$1.2{\cdot}10^{-4}$	$2.5{\cdot}10^{-5}$	$1.2{\cdot}10^{-4}$	$2.1{\cdot}10^{-5}$	$1.2{\cdot}10^{-4}$	$1.2{\cdot}10^{-5}$	$2.8{\cdot}10^{-5}$	$9.4{\cdot}10^{-6}$	$2.6{\cdot}10^{-5}$	$4.8{\cdot}10^{-6}$	$6.6{\cdot}10^{-6}$	$4.7{\cdot}10^{-6}$	$6.2{\cdot}10^{-6}$
σ^2	0.0024	0.0522	0.0008	0.0015	0.0188	0.0523	0.0003	0.0010	0.0087	0.0306	0.0001	0.0007	0.0039	0.0118
ρ	1.5400	2.1400	0.4882	0.4804	0.6667	1.6004	0.2203	0.1003	0.3001	0.9397	0.1450	0.0781	0.1223	0.4060
γ	-0.0182	-	0.0034	0.0138	-	-	0.0027	0.0055	-	-	0.0022	0.0022	-	-
σ^2_{VG}	0.0069	-	0.0011	0.0070	-	-	0.0007	0.0027	-	-	0.0003	0.0009	-	-
ν	2.4900	-	0.4601	2.0040	-	-	0.3944	0.9667	-	-	0.1808	0.3356	-	-
σ_{LS}	-	0.0308	-	-	0.0051	0.0202	-	-	0.0032	0.0097	-	-	0.0015	0.0043
α	-	1.7115	-	-	0.3206	1.8015	-	-	0.2015	0.8882	-	-	0.1111	0.3749

Notes: (i), The models are DGP 2.9, CKLS-VG model and DGP 2.10, CKLS-LS models in (2.58) and (2.59) respectively. (ii), The sampling frequency is daily with $\Delta = 1/252$. (iii), The number of iterations is 1000.

stable variables to our target stable variables, which have a drift of zero and dispersion of $\sigma_{LS}\Delta^{1/\alpha}$.

The results for DGPs 2.9-2.10 are reported in Table 2.5. It can be observed that the proposed LEL estimator can accurately estimate the parameters of both CKLS-VG and CKLS-LS models in general. For CKLS-VG, the empirical biases in percentages for jump parameters γ, σ_{VG}^2, and ν are 18.68%, 15.94% and 18.48% respectively when n is only 1000 while those for the jump parameters σ_{LS} and α are 16.56% and 18.73% respectively for n=1000. Hence, given that the Levy jumps happen frequently, my LEL method is able to identify the jump parameters accurately. Of course, similar to DGPs 2.1-2.2, large biases are still present for the mean-reverting parameters.

2.4 Exchange Rate Dynamics with Levy Jumps

In this section, I shall conduct an empirical study of Levy jumps in exchange rates. One difficulty with such works is that econometric inferences are rather complicated. For instance, except for special cases, the probability densities of most Levy processes are not available in analytic form, and for certain processes, not all moments exist. The main approaches employed by the existing literature include numerical Fourier transform-based likelihood methods in Carr, Geman, Madan, and Yor (2002), Bayesian methods via MCMC simulations in Li, Wells, and Yu (2008, 2009), efficient method of moments in Kretschmer and Pigorsch (2004), and Kalman filer methods in Carr and Wu (2004b). Most of these procedures are computationally demanding. In contrast, the proposed infinitesimal operator-based LEL approach is very convenient to implement.

2.4.1 Jump-Diffusion Models Driven by Levy Processes

I shall first introduce general Levy processes following Li, Wells and Yu (2008, 2009) and then discuss the Levy driven jump-diffusion models employed in the current work. Suppose X_t is a scalar Levy process adapted to $\mathfrak{F}_t$. Then the sample path of X_t is Cadlag and $X_s - X_t$ is independent of $\mathfrak{F}_t$ and identically distributed with X_{s-t} for $0 \leq t < s$. Clearly, both Brownian motion and compound Poisson process are special cases of Levy processes. While the probability densities of Levy processes are generally not available in closed form, their characteristic functions $\phi_{X_t}(u)$ can be explicitly obtained as follows: $\phi_{X_t}(u) = E\left[e^{iuX_t}\right] = e^{-t\psi_x(u)}$, $t \geq 0$, where $\psi_x(\cdot)$ is called the characteristic exponent and satisfies the following Levy-Khintchine formula (see Bertoin, 1996, p.12):

$$\psi_x(u) \equiv (-i\mu)\,u + \frac{\sigma_L^2}{2}u^2 + \int_{\mathbb{R}_0}\left(1 - e^{iux} + iux1_{|x|\leq 1}\right)\pi(dx),$$

$u \in \mathbb{R}$, $\mu \in \mathbb{R}$, $\sigma_L \in \mathbb{R}^+$ and π is a measure on $\mathbb{R}_0 = \mathbb{R}\backslash\{0\}$ with $\int_{\mathbb{R}_0} \min\left(1, x^2\right)\pi(dx) < \infty$ which implies that the process has finite quadratic variation. The triplet $\left(\mu, \sigma_L^2, \pi(\cdot)\right)$, usually referred to as Levy characteristic of an infinitely divisible distribution (Bertoin, 1996), characterizes the dynamic probability law of the Levy process. The Levy measure $\pi(dx)$ determines the jump structure of the process, with the interpretation that $\pi(E)$, for any subset $E \subset \mathbb{R}$, is the rate at which the process takes jumps of size $x \in E$. In other words, $\pi(E)$ measures the number of jumps whose jump sizes fall in E per unit of time. In this sense, the Levy measure $\pi(dx)$ captures the jump intensity and jump sizes in an integral way.

A pure jump Levy process can exhibit rich jump features depending on its Levy measure $\pi(\cdot)$. The so-called finite-activity jump processes only generate a finite number of jumps within any finite time interval and therefore $\pi(\cdot)$ needs to be integrable, i.e.,

$$\int_{\mathbb{R}_0} \pi\,(dx) = \lambda < \infty \tag{2.54}$$

The classical example of such a finite-activity jump process is the compound Poisson process of Merton (1976). Different from the finite-activity jump process, an infinite-activity jump process can generate an infinite number of jumps within any finite time interval. Consequently, the integral of the Levy measure in (2.54) is no longer finite. Furthermore, within the infinite-activity class, the sample path of the jump process can exhibit either finite or infinite variation, implying that the accumulated absolute distance traveled by the process over any finite time interval is finite or infinite, respectively.

In the empirical studies, I choose a relatively parsimonious VG model as a representative of the infinite-activity but finite-variation jump model, similar to Li, Wells and Yu (2008, 2009). The VG process can be obtained by subordinating an arithmetic Brownian motion with drift γ and variance σ by an independent gamma process with unit mean rate and variance rate ν, G_t^ν. That is,

$$X_{VG}\,(t|\sigma,\gamma,\nu) = \gamma G_t^\nu + \sigma W\,(G_t^\nu)$$

where $W\,(t)$ is a standard Brownian motion and is independent of G_t^ν. The Levy measure of the VG process is

$$\pi_{VG}\,(dx) = \frac{A_\pm^2 \exp\left(-\frac{A_\pm}{B_\pm}|x|\right)}{B_\pm |x|}\,(dx) \tag{2.55}$$

where $A_\pm = \frac{1}{\nu}\left(\sqrt{\frac{\gamma^2\nu^2}{4}+\frac{\sigma^2\nu}{2}} \pm \frac{\gamma\nu}{2}\right)$ and $B_\pm = A_\pm^2\nu$. The parameters with plus subscripts apply to positive jumps and those with minus subscripts to negative jumps. If $\gamma = 0$, the jump structure is symmetric around zero and the subscripts can be dropped. Note that when the jump size approaches zero, the arrival rate is close to infinity. Thus, this infinite-activity process incorporates (possibly) infinitely many small jumps. In addition the Levy measure of an infinite-activity jump process is singular at a zero jump size.

Another example of an infinite-activity jump model is the Levy α-stable process. In this process, the jump size follows a so-called α-stable distribution denoted as $S_\alpha(\beta, \delta, \gamma)$ with a tail index $\alpha \in (0, 2]$, a skew parameter $\beta \in [-1, 1]$, a scale parameter $\delta \geq 0$, and a location parameter $\gamma \in \mathbb{R}$. The parameter α controls the shape while β determines the skewness of the distribution. The stable densities are supported on either $\mathbb{R}$ or $\mathbb{R}^+$, with the latter case occurring only when $\alpha < 1$ and $\beta = \pm 1$. The characteristic function of an α-stable distribution S is given by

$$E\left[e^{iuS}\right] = \begin{cases} \exp\left(-\delta^\alpha |u|^\alpha \left[1 - i\beta\left(\tan\frac{\pi\alpha}{2}\right) sign(u)\right] + i\gamma u\right) & \alpha \neq 1 \\ \exp\left(-\delta |u| \left[1 + i\beta\frac{2}{\pi} sign(u) \ln |u|\right] + i\gamma u\right) & \alpha = 1 \end{cases} \tag{2.56}$$

For a standardized α-stable distribution, denoted as $S_\alpha(\beta, 1, 0)$, $\delta = 1$ and $\gamma = 0$.

All α-stable processes are built by a fundamental process called α-stable motion. By definition, a process X_t is an α-stable motion if (i) $X_t = 0$ a.s., (ii) X_t has independent increments, and (iii) the increment $X_t - X_s$ $(t > s)$ follows an α-stable distribution $S_\alpha\left(\beta, (t-s)^{1/\alpha}, 0\right)$. The role played by the α-stable motion for the α-stable process is similar to that of Brownian motion for diffusion processes. Among α-stable processes, we choose the LS process of Carr and Wu (2003) in our empirical studies. We obtain this process by multiplying an α-stable motion by a constant σ. Following Carr and Wu (2003), we set $\beta = -1$ to achieve finite moments and negative skewness in the process density, a feature that cannot be captured by either a Brownian motion or a symmetric Levy α-stable motion. We further restrict $\alpha \in (1, 2)$ in order that the process be supported on the whole real line. The resulting α-stable process defined in this way is a Levy process with infinite activity and infinite variation and has the Levy measure

$$\pi_{LS}(x) = c_{\pm} |x|^{-1-\alpha} dx \tag{2.57}$$

where $c_{-} = \frac{-\sigma^{\alpha} \sec \frac{\pi\alpha}{2}}{\Gamma(-\alpha)}$. The parameters $c_{\pm}$ control both the scale and asymmetry of the process. In the LS model, c_{+} is set to zero so that only negative jumps are allowed in the Levy measure. However, in addition to the pure jump part characterized by the Levy measure $\pi_{LS}(dx)$, the LS process has also a deterministic drift part which compensates the negative jumps to make the whole process a martingale. For infinite-variation jumps, the compensation is strong enough to support the whole real line as the admissible domain of the LS process that only accommodates negative jumps. As a result, the LS process has an α-stable distribution with infinite p-th moment for $p > \alpha$.

To examine the significance of Levy jumps in capturing the observed frequent but small potential jumps in the exchange rates, I consider the following CKLS-type jump-diffusion models driven by Levy processes and by compound Poisson process.

- CKLS Model with VG Jumps (CKLS-VG):

$$dX_t = \kappa(\overline{\alpha} - X_t) dt + \sigma X_t^{\rho} dW_t + dJ_t^{VG} \tag{2.58}$$

  where J_t^{VG}, the VG process with Levy measure in (2.55), is independent of W_t. In fact, $J_t^{VG} = \gamma G_t^{\nu} + \sigma W(G_t^{\nu})$ with the gamma process G_t^{ν}, having unit mean rate and variance rate ν, independent of W_t.

- CKLS Model with LS Jumps (CKLS-LS):

$$dX_t = \kappa(\overline{\alpha} - r_t) dt + \sigma X_t^{\rho} dW_t + dJ_t^{LS} \tag{2.59}$$

where J_t^{LS}, the LS process with Levy measure in (2.57), is independent of W_t. The increment of LS process J_t^{LS} follows an α-stable distribution $S_\alpha\left(-1, \Delta^{1/\alpha}, 0\right)$ with shape parameter α, skewness parameter -1, zero drift and scale parameter $\sigma_{LS}\Delta^{1/\alpha}$.

- CKLS Model with compound Poisson Jumps (CKLS-P):

$$dX_t = \kappa\left(\overline{\alpha} - X_t\right)dt + \sigma X_t^{\rho} dW_t + dJ_t^P \tag{2.60}$$

where $J_t^P = \xi_t N_t$ is independent of W_t, N_t is a compound Poisson process with jump intensity λ, and $\xi_t \sim N\left(\mu_y, \sigma_y^2\right)$ is the jump size.

2.4.2 Estimation Results for Exchange Rate Dynamics

It can be seen that all three models in (2.58)-(2.60) are special cases of Model (2.1). Hence the LEL approach can be applied for the estimation. The data I use are daily Euro/Dollar and Yen/Dollar rates from June 25, 2000 to June 25, 2010 and June 25, 1990 to June 25, 2010 respectively (note that the recent financial crisis starting in 2008 is included), obtained from Datastream. Euro and Yen are two of the most important currencies in the world in addition to the U.S. dollar. The launch of the new currency Euro has created the world's second largest single currency area after the United States and understanding the evolution of the Euro/Dollar exchange rates will be important to many outstanding issues in international economics and finance. The Japanese economy has been in a prolonged recession for more than a decade, and hence the Yen/Dollar rate might have very different time-series properties than that of the Euro/Dollar rate. The starting time for Euro/Dollar sample data, i.e., June, 2000, is chosen to allow the market to stabilize after the introduction of the Euro as a new currency on

Table 2.6: **Summary Statistics for Euro/Dollar and Yen/Dollar Rates**

	Mean	Standard Deviation	Skewness	Kurtosis	Min	Max
Euro/Dollar	$-1.0627 \cdot 10^{-4}$	0.0064	-0.1174	5.2572	-0.0425	0.0256
Yen/Dollar	$-1.0583 \cdot 10^{-4}$	0.0070	-0.5541	7.4281	-0.0549	0.0324

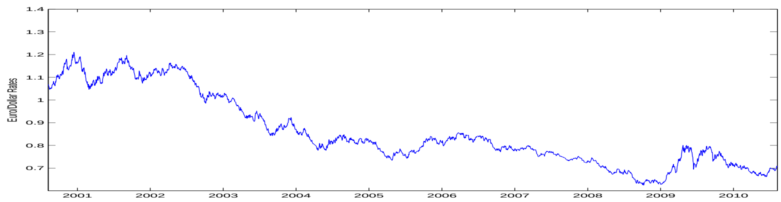
Notes: This table reports the summary statistics for log differences of Euro/Dollar and Yen/Dollar rates at daily frequency from June 25, 2000 to June 25 2010 and from June 25, 1990 to June 25, 2010 respectively.

January 1, 1999. Following Diebold et al. (1999) and Hong, Li, and Zhao (2007), I calculate the log difference of exchange rate levels at consecutive time points. Table 2.6 provides the summary statistics and Figure 2.1 the time-series plots. It can be seen that both Euro/Dollar and Yen/Dollar rates are negatively skewed and have fatter tails than the normal distribution. In addition, the latter has slightly higher kurtosis than the former.

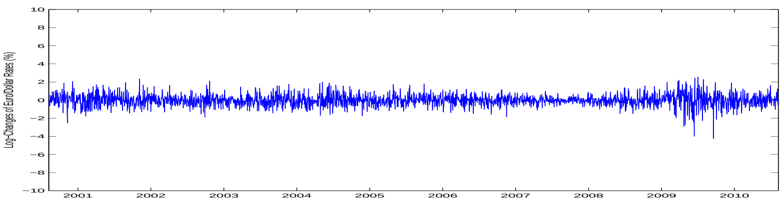
The estimation results for the three models in (2.58)-(2.60) using both Euro/Dollar and Yen/Dollar rates are reported in Table 2.7. For all three models, the estimated mean-reverting speed, ranging from 0.4458 to 1.0252, suggests that both Euro/Dollar and Yen/Dollar returns revert to their mean level very quickly. The estimated long-run mean is very similar across the three models, with values around the mean in the summary statistics in Table V, i.e., 0.01%. For the jump terms of CKLS-P, the estimated intensities, 0.0106 for Euro and 0.0128 for Yen, imply that both series jump around 3 times a year. But the average jump sizes, although both negative, are bigger for Euro (-0.0210) than for Yen (-0.0280). For the jump terms of both CKLS-VG and CKLS-LS, the estimated

Figure 2.1: **Time Series of Yen/Dollar and Euro/Dollar Rates**

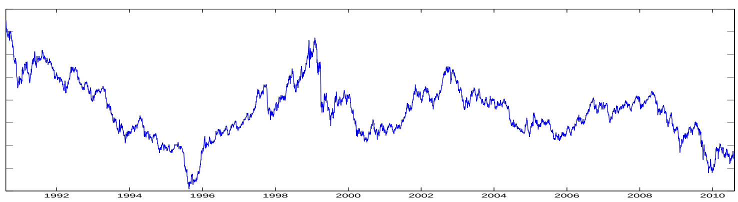
This figure plots levels and log-changes of daily Euro/Dollar and Yen/Dollar rates from June 25, 2000 and June 25, 1990 to June 25, 2010 respectively.



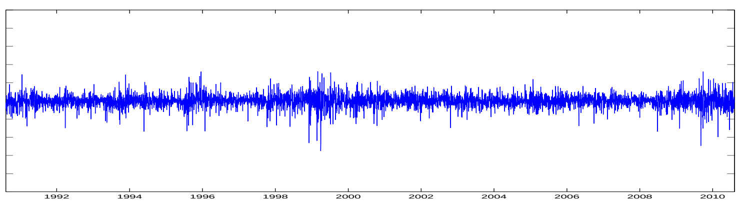
(a) Level of Euro/Dollar Rates



(b) Log Changes of Euro/Dollar Rates



(c) Level of Yen/Dollar Rates



(d) Log Changes of Yen/Dollar Rates

Table 2.7: **Parameter Estimates for Levy Jump Models of Exchange Rates**

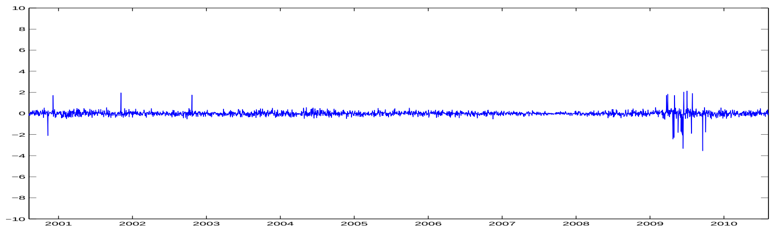
	Euro/Dollar						Yen/Dollar					
	CKLS-P		CKLS-VG		CKLS-LS		CKLS-P		CKLS-VG		CKLS-LS	
	Est.	SE	Est.	SE	Est.	SE	Est.	SE	Est.	SE	Est.	SE
κ	1.0252	(0.0167)	0.8426	(0.5576)	0.8223	(0.0836)	0.6764	(0.0842)	0.8951	(0.0058)	0.4458	(0.0808)
$\overline{\alpha}$	$1.1{\cdot}10^{-4}$	$(1.2{\cdot}10^{-5})$	$1.5{\cdot}10^{-4}$	$(1.6{\cdot}10^{-5})$	$1.2{\cdot}10^{-4}$	$(3.0{\cdot}10^{-5})$	$1.2{\cdot}10^{-4}$	$(2.1{\cdot}10^{-5})$	$1.3{\cdot}10^{-4}$	$(7.1{\cdot}10^{-5})$	$1.1{\cdot}10^{-4}$	$(2.4{\cdot}10^{-5})$
σ^2	0.0088	(0.0024)	0.0024	(0.0019)	0.0522	(0.0452)	0.0058	(0.0012)	0.0075	(0.0008)	0.0822	(0.0104)
ρ	1.2110	(0.6906)	1.5400	(0.1321)	2.1400	(0.4427)	0.6110	(0.0047)	2.5400	(0.1923)	3.1400	(1.4569)
λ	0.0106	(0.0054)	-	-	-	-	0.0128	(0.0030)	-	-	-	-
μ_y	-0.0210	(0.0025)	-	-	-	-	-0.0280	(0.0149)	-	-	-	-
σ_y^2	0.0068	(0.0022)	-	-	-	-	0.0024	(0.0005)	-	-	-	-
γ	-	-	-0.0182	(0.0047)	-	-	-	-	-0.0266	(0.0036)	-	-
σ_{VG}^2	-	-	0.0069	(0.0028)	-	-	-	-	0.0068	(0.0007)	-	-
ν	-	-	2.4900	(0.0056)	-	-	-	-	4.6900	(0.0011)	-	-
σ_{LS}	-	-	-	-	0.0308	(0.0025)	-	-	-	-	0.0086	(0.0009)
α	-	-	-	-	1.7115	(0.9275)	-	-	-	-	1.2653	(0.7594)

Notes: This table reports the LEL estimates (with standard errors in the parentheses) of the CKLS-P, CKLS-VG, and CKLS-LS models, using log differences of Euro/Dollar and Yen/Dollar rates at daily frequency from June 25, 2000 and June 25, 1990 to June 25, 2010 respectively.

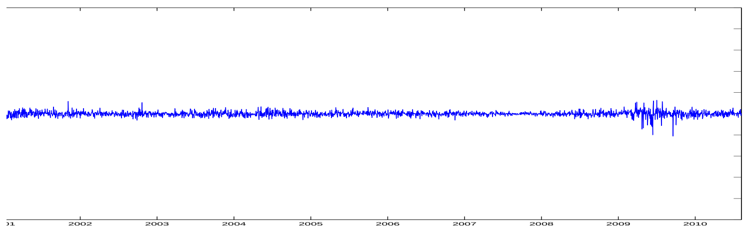
parameters are all significantly different from zero, implying the importance of Levy jumps in both series. Furthermore, the estimated jump parameters (γ and ν) for CKLS-VG are of larger magnitude for Yen than for Euro series, while those for CKLS-LS (σ_{LS} and α) have the reversed scenario.

Finally, motivated by Johannes (2004), I examine the filtered jump variables for all three models by estimating the filtering distribution $E[J_{t+\Delta}|X_{t+\Delta}, X_t, \widehat{\theta}_{LEL}]$ where $\widehat{\theta}_{LEL}$ is the parameter estimator. Since no latent variables are involved, it is straightforward to compute the filtering distribution by iteratively sampling from $p[J_{t+\Delta}|X_{t+\Delta}, X_t, \widehat{\theta}_{LEL}]$, which is a standard distribution. The algorithm produces a sequence $\{\{J^g_{t+\Delta}\}^{n\Delta}_{t=\Delta}\}^G_{g=1}$ where G is the number of iterations and we choose $G = 5000$ and discard the first 2000 iterations as a burn-in period. Figure 2.2 presents the filtered jump variables for Euro/Dollar and Figure 2.3 for Yen/Dollar. Observe that significant jumps happen during 2001-2003 and 2009-2010 for Euro and during 1994-1996, 1999-2000 and 2009-2010 for Yen. However, Figure 2.2, (a) and (b) and Figure 2.3, (a) and (b), which present the filtered jump variables for CKLS-VG and CKLS-LS, show that there are many small but frequent jumps during the sample period for both Euro and Yen. These small jumps happen so frequently that they are most likely induced by normal market information flows such as those related to transactions rather than by big economic announcements as in Johannes (2004). Figure 2.2, (c) and Figure 2.3 (c) show that although the filtered Levy jumps are similar, there are still some differences between VG- and LS- driven models.

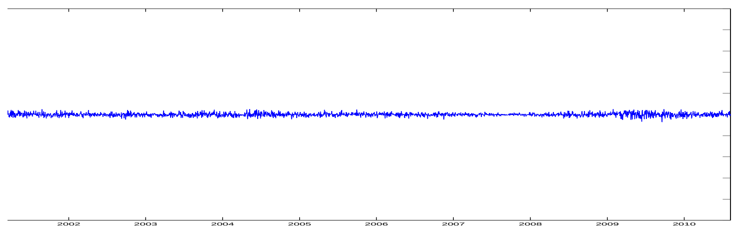
Figure 2.2: **Filtered Jump Variables for Euro/Dollar Rates**



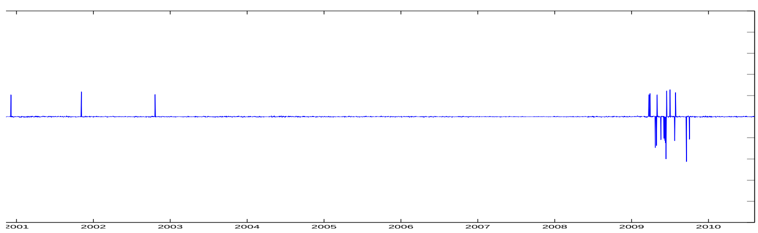
(a) CKLS-VG: Estimated Jumps



(b) CKLS-LS: Estimated Jumps

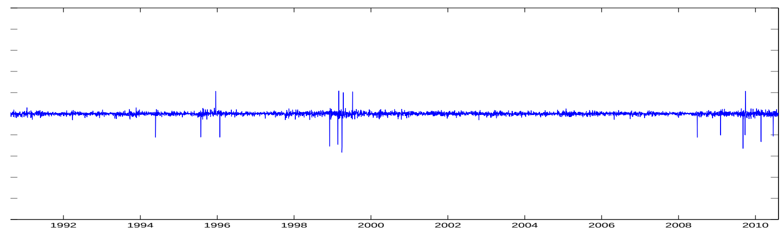


(c) Differences in Estimated Jumps: CKLS-VG and CKLS-LS

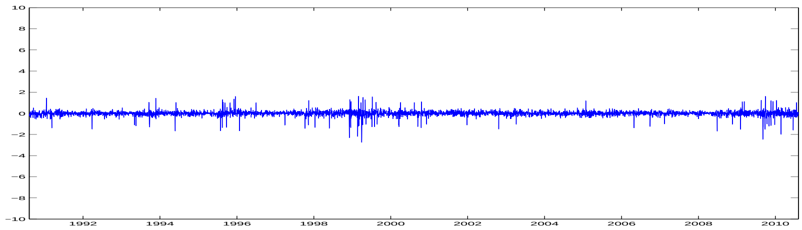


(d) CKLS-P: Estimated Jumps

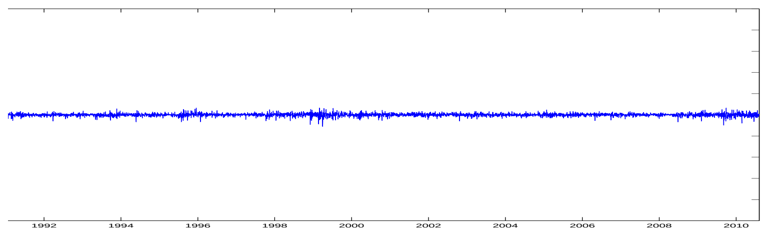
Figure 2.3: **Filtered Jump Variables for Yen/Dollar Rates**



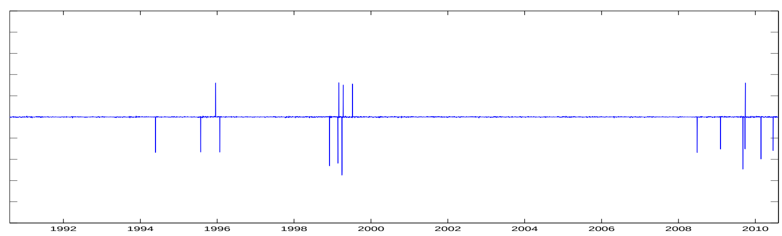
(a) CKLS-VG: Estimated Jumps



(b) CKLS-LS: Estimated Jumps



(c) Differences in Estimated Jumps: CKLS-VG and CKLS-LS



(d) CKLS-P: Estimated Jumps

2.5 Conclusion

A local empirical likelihood estimator is proposed for a general continuous-time multivariate Markov model. Most popular continuous-time finance models are covered as special cases, including diffusion, jump-diffusion and Levy jump-diffusion models. Avoiding the inconvenient transition density, my method is based on the infinitesimal operator, which is available in closed form. This renders my estimator particularly convenient for multivariate cases. The proposed estimator, via a local empirical likelihood approach, is asymptotically efficient and involves no need to estimate the optimal instruments. Simulation studies show that its performance is comparable to the (exact, approximated, and simulated) MLE. An empirical application of Levy jumps in exchange rate dynamics using the Euro/Dollar and Yen/Dollar data is conducted by the proposed method.

Note that this framework does not cover stochastic volatility (SV) models, which are very popular for derivative pricing due to their ability to capture the empirical features of the joint time-series behavior of the underlying asset and its derivatives prices (Stein and Stein (1991); Heston (1993)). However, the proposed LEL estimator can be readily extended for such models as long as an appropriate proxy is available in place of the unobservable volatility. This is similar to the analysis in Ait-Sahalia and Kimmel (2007), which extends the AMLE of Ait-Sahalia (2002, 2008) to SV models by replacing the unobservable volatility process using either Black-Scholes implied volatility of an at-the-money short-maturity option or a proxy from inverting the observed option prices (see Pan (2002)). Another strategy is to use the realized volatility estimated from high-frequency data, like Kanaya and Kristensen (2009) and Bandi and Reno (2008).

Another valuable extension, as discussed in Section 1, is to estimate the multifactor affine jump-diffusion term structure models which do not perfectly fit into the framework of the current paper. The reason is that these affine term structure models contain general unobservable state variables as risk factors while the framework in (2.1) is a general continuous time Markov process with no latent variables. Chapter 3 takes up this challenge, adapting the infinitesimal operator-based LEL estimator proposed here to estimate multifactor AJD term structure models by extracting the latent state variables from the observed bond yields and then applying the estimator to conduct an empirical study of three-factor AJD term structure models for the LIBOR-Swap rates.

Finally, it is interesting to investigate whether the proposed infinitesimal operator-based method can be extended to cases in which the time separating successive observations of a continuous-time model is random (see Ait-Sahalia and Mykland (2003, Figure 1, p. 484) for an illustration of the situation). Ait-Sahalia and Mykland (2003) study the effect of sampling randomness when estimating a continuous-time model by comparing the properties of three likelihood-based estimators. Duffie and Glynn (2004) develop a family of generalized method-of-moments estimators of a continuous-time Markov process observed at random time intervals. However, their method is specific to the type of random sampling assumed in their study (in particular, a Poisson sampling occurring at an arrival intensity) and does not allow for the important special case where sampling occurs at fixed time intervals. Since the conditional moment restriction in this study is derived as an implication of the martingale property of the transformed processes which does not depend on any specific sample schemes, the LEL estimator has the potential to be generalized for such random sampling cases.

CHAPTER 3

EXPECTATION PUZZLES, TIME-VARYING CONDITIONAL VOLATILITY AND JUMPS IN AFFINE TERM STRUCTURE MODELS

3.1 Expectation Puzzles and Time-Varying Conditional Volatility

In this section, I shall discuss first the "expectation hypothesis," the empirical pattern of deviations from which characterizes time variations of the risk premium (Campbell and Shiller, 1991; Dai and Singleton, 2002). After introducing LIBOR-Swap yields data, the empirical pattern of violations of the "expectation hypothesis" is documented. Finally, the term structure of yields volatility is also provided to illustrate time variations in the conditional volatility of LIBOR-Swap rates.

3.1.1 The Expectation Hypothesis

The "expectation hypothesis" is probably the oldest and most highly regarded classical theory of the term structure of interest rates (Fisher, 1896; Lutz, 1940). It is a theory that links bond returns, yields and forward rates of a wide range of terms of maturity. Jarrow (2009) summarizes three formulations of the "expectation hypothesis": the local expectations (**LE**) hypothesis, the return-to-maturity expectations hypothesis, and the unbiased expectations hypothesis. Each of these formulations can be characterized by a formula for calculating the zero-coupon bond's price. Like Campbell and Shiller (1991), Backus et al. (2001), and

Dai and Singleton (2002), I focus mainly on the **LE** form and in fact its empirical implication in terms of a predictive regression.

To fix some notations, we use P_t^τ to denote the time-t price of a zero-coupon bond with τ periods to mature, y_t^τ ($\equiv -\ln P_t^\tau/\tau$) to denote its corresponding continuously compounded yield, and $r_t \equiv y_t^1$ to denote the short-term interest rate. We define one-period expected excess holding period returns as

$$e_t^\tau \equiv E_t\left[\ln\left(P_{t+1}^{\tau-1}/P_t^\tau\right)\right] - r_t \tag{3.1}$$

and then by definition we have

$$e_t^\tau / (\tau - 1) + E_t\left[y_{t+1}^{\tau-1} - y_t^\tau\right] = (y_t^\tau - r_t) / (\tau - 1) \tag{3.2}$$

where $E_t\left[y_{t+1}^{\tau-1} - y_t^\tau\right]$ is the expected one-period yield change and $(y_t^\tau - r_t) / (\tau - 1)$ is the average yield spread (also called the slope of the term structure).

From Jarrow (2009), the **LE** form of the "expectation hypothesis" implies that $e_t^\tau \equiv 0$, delivering the equality of the expected yield change and yield spread. This can serve as a test of the "expectation hypothesis." To accommodate other effects such as trading costs, most studies of the "expectation hypothesis" tests generalize the **LE** hypothesis to $e_t^\tau \equiv constant$, which does not depend on time t but may not be equal to zero. Combined with (3.2), this brings us the so-called "yield regression"[1] used in Campbell and Shiller (1991), Backus et al. (2001), and Dai and Singleton (2002):

$$y_{t+1}^{\tau-1)} - y_t^\tau = constant + \phi_{\tau T}\left(\frac{y_t^\tau - r_t}{\tau - 1}\right) + residual \tag{3.3}$$

[1]Similarly, other measures of risk premiums can be defined such as the yield term premiums ($c_t^\tau \equiv y_t^\tau - \frac{1}{\tau}\sum_{i=0}^{\tau-1} E_t\left[r_{t+i}\right]$) and forward term premiums ($p_t^\tau \equiv f_t^\tau - E_t\left[r_{t+i}\right]$, where $f_t^\tau \equiv -\ln\left(P_t^{\tau+1}/P_t^\tau\right)$ is the forward rate), which lead to a range of empirical predictive regressions for testing the "expectation hypothesis"; for details, see Fama (1984a, b, 2006), Fama and Bliss (1987), Campbell and Shiller (1991), Backus et al. (2001), Bekaert et al. (1997, 2001), Dai and Singleton (2002), and Stambaugh (1988).

If the "expectation hypothesis" holds in the generalized sense of $e_t^\tau \equiv constant$, we should have $\phi_{\tau T} = 1$ for all τ. On the contrary, any deviation from the benchmark of $\phi_{\tau T} = 1$ for all τ implies that the expected excess return, that is, the risk premium, is time-varying and the empirical deviation pattern captures in essence the time variation in the risk premium, which is termed **LPY (i)** (linear projection of yields) in Dai and Singleton (2002).

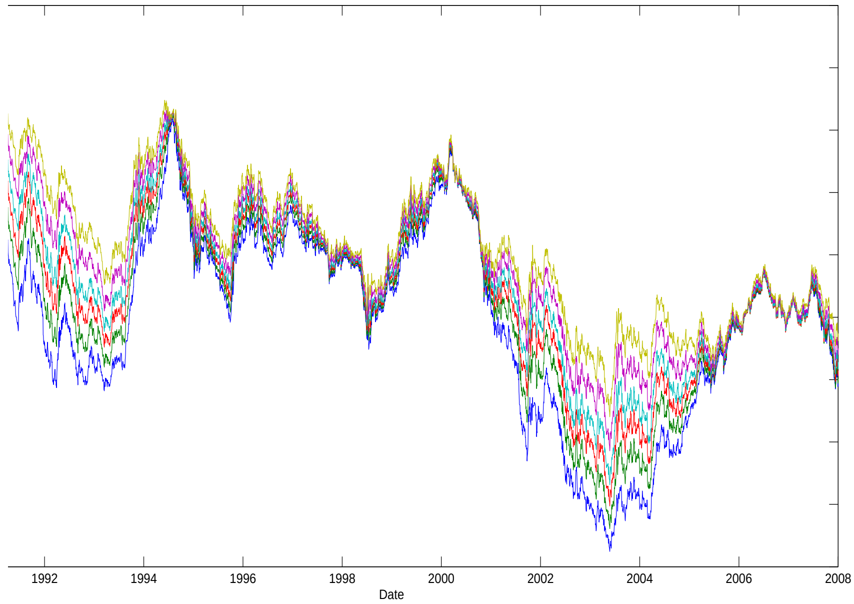
3.1.2 Data

Whereas most empirical studies of the "expectation hypothesis" have focused on monthly U.S. Treasury yield data (Fama and Bliss, 1987; Campbell and Shiller, 1991; Backus et al., 2001; Dai and Singleton, 2002), I follow Dai and Singleton (2000; 2003) and Piazzesi (2005) in choosing LIBOR-Swap rates in the empirical analysis with the main motivation that high frequency daily data are readily available and more relevant for studying jumps. Another advantage of swap rates is that they are truly constant maturity yields, rendering interpolation unnecessary. Of course, swap contracts have been traded only since the end of the 1980s and hence the periods that include the oil price shocks of the early 1970s and the monetary experiment of the early 1980s are not covered. See Dai and Singleton (2000), Johannes and Sundaresan (2007), and Piazzesi (2010) for further discussion of swap rates.

The data contain daily LIBOR-Swap rates with 3-month, 6-month, 9-month, 2-year, 3-year, 4,-year, 5-year, 7-year and 10-year maturities, obtained from Datastream. The sample period is from August 13, 1990 to December 31, 2008 with a total of 4757 observations, determined in part by the unavailability of

Figure 3.1: **Time Series of LIBOR-Swap Rates**

This figure plots daily LIBOR-Swap rates from August 13, 1990 to December 31, 2008 with a total of 4757 observations. The rates plotted have, from the lowest to the highest line with occasional cross-overs when the yield curves are inverted, 3-month, 6-month, 9-month, 2-year, 3-year, 4-year, 5-year, 7-year and 10-year maturities respectively.



reliable swap data prior to 1987. Note that the recent financial crisis of 2008 is covered. Figure 3.1 provides a time series plot of these LIBOR-Swap yields.

Table 3.1 reports the summary statistics of the levels of and changes in LIBOR-Swap yields. We can see that long-term yields tend to be higher than short-term yields and hence the yield curve is upward sloping. On average, all yields exhibit negative changes, which is consistent with the declining interest rates during the sample period. The standard deviation of yield changes

Table 3.1: **Summary Statistics of LIBOR-Swap Rates**

This table reports the summary statistics of daily LIBOR-Swap rates with maturities of 3-month, 6-month, 9-month, 2-year, 3-year, 4,-year, 5-year, 7-year and 10-year from August 13, 1990 to December 31, 2008. Panel A and B provide summary statistics of the levels and changes of LIBOR-Swap yields respectively. Panel C provides the results of principal component analysis of LIBOR-Swap yields. The entries represent the percentages of the variations of the levels and changes of LIBOR-Swap yields explained by each of their first six principal components.

				Panel A					
	3-m	6-m	9-m	2-y	3-y	4-y	5-y	7-y	10-y
Mean(%)	4.413	4.505	4.595	5.022	5.304	5.521	5.692	5.936	6.169
Std. Dev(%)	1.751	1.742	1.739	1.631	1.536	1.470	1.425	1.366	1.311
Skewness	-0.367	-0.391	-0.385	-0.239	-0.113	0.003	0.104	0.263	0.357
Kurtosis	2.210	2.242	2.267	2.407	2.471	2.485	2.473	2.496	2.474
AutoCorr	0.998	0.998	0.998	0.998	0.998	0.998	0.997	0.997	0.997
				Panel B					
Mean(%)	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001
Std. Dev(%)	0.027	0.031	0.034	0.017	0.015	0.014	0.013	0.012	0.011
Skewness	-0.267	-0.430	-0.171	0.089	0.137	0.181	0.185	0.160	0.106
Kurtosis	163.365	247.155	237.738	6.282	6.537	5.700	5.697	5.686	5.793
AutoCorr	0.003	-0.104	-0.113	0.042	0.019	0.025	0.019	0.021	0.028
				Panel C					
	1	2	3	4	5	6	7	8	9
Level (%)	92.50	7.11	0.34	0.03	0.01	0.01	0.00	0.00	0.00
Change(%)	73.28	20.99	2.61	1.44	0.62	0.38	0.30	0.20	0.18

increases when the maturity rises from 3 months to 9 months and then decreases with maturities longer than 9 months. Changes in short-term yields exhibit higher kurtosis and are more negatively skewed than changes in long-term yields. Yield levels are very persistent, with first-order autoregressive coefficients close to one. In contrast, yield changes are much less persistent with the first-order autoregressive coefficients ranging from -0.1134 to 0.0424. Principal component analysis shows that, as is the case with U.S. Treasury data (Litterman and Scheinkman, 1997), the first three principal components can explain more than 99.9% and 96.9% of the variations in the levels of and changes in yields, respectively. This confirms the claim in Dai and Singleton (2000) that some of the basic distributional characteristics of Treasury and LIBOR-Swap yields are similar even though the institutional structures are different for the two markets, which justifies studying the "expectation hypothesis" using swap yields.

3.1.3 Expectation Puzzles for LIBOR-Swap Yields

To test the "expectation hypothesis" using the "yield regression" in (3.3), we note that zero-coupon bond yields need to be used due to the dependence of the **LE** hypothesis on y_t^τ. However, the swap yield is equivalent to a par−bond yield with the coupon rate equal to the swap yield.[2,3] To be consistent with the literature that examines term structure dynamics using swap rates directly (Dai and Singleton, 2000; Piazzesi, 2005), I treat swap rates as approximations of the corresponding zero-coupon yields. Appendix B describes the differences between swap rates and constructed zero-coupon yields, showing that errors in

[2] I am very grateful to Pamela Moulton for pointing out the differences to me.

[3] Here I follow Collin-Dufresne and Goldstein (2002) and Li and Zhao (2006) to assume that the quoted swap rate is equivalent to a par-bond rate for an issuer with LIBOR-credit quality. See Johannes and Sundaresan (2007) and Piazzesi (2010) for further discussion of this assumption.

Table 3.2: **Yield Regression Using LIBOR-Swap Yields**

This table reports the results of "yield regression" in (2.3) using daily LIBOR-Swap rates with maturities as indicated in the table from August 13, 1990 to December 31, 2008. The 3-month LIBOR rate is used as the spot rate r_t. In the "s.e." row are the Newey-West standard errors of $\phi_{\tau T}$.

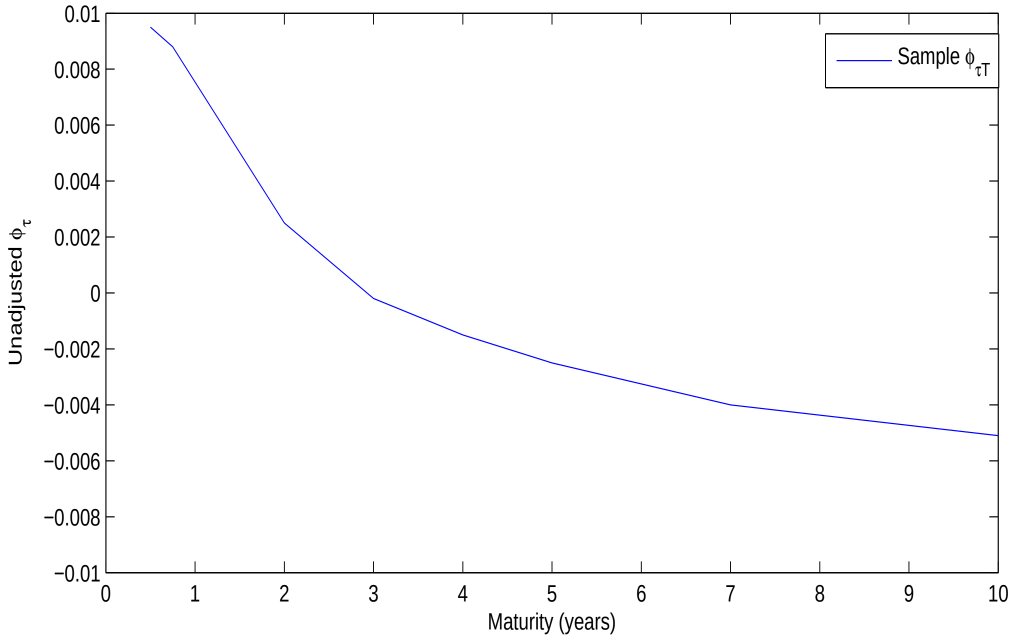
	$y_{t+1}^{(\tau-1)} - y_t^{\tau}$ = constant+$\phi_{\tau T}\,(y_t^{\tau} - r_t)\,/\,(\tau-1)$ + $residual$							
	6-m	9-m	2-y	3-y	4-y	5-y	7-y	10-y
ϕ_{nT}	0.0095	0.0088	0.0025	-0.0002	-0.0015	-0.0025	-0.0040	-0.0051
$s.e.$	0.0027	0.0026	0.0031	0.0038	0.0043	0.0008	0.0038	0.0022

these approximations are very small and do not affect the results when testing the "expectation hypothesis."

Estimates of slope coefficients $\phi_{\tau T}$ in the "yield regression" of (3.3) using LIBOR-Swap rates are reported in Table 3.2 and Figure 3.2 , with the 3-month LIBOR rate treated as the spot rate r_t. We can see that the estimated coefficients $\phi_{\tau T}$ are close to zero overall for all τ. They decrease with larger maturity τ, changing from slightly positive values for rates with maturities less than 2 years to slightly negative values for rates with maturities longer than 2 years. Although the standard errors are not very small, the estimated coefficients are still significantly different from one by conventional statistical standards. This empirical pattern of violations of the "expectation hypothesis" for LIBOR-Swap yields is similar to that for U.S. Treasury yields in terms of the increasingly negative regression coefficients with longer maturity τ (Dai and Singleton, 2002). They differ, however, insofar as the former has much smaller regression coef-

Figure 3.2: **Estimated Coefficients in the "Yield Regression"**

This figure plots coefficients estimates $\phi_{\tau T}$ in the "yield regression" $y_{t+1}^{(\tau-1)} - y_t^{\tau}$ = constant + $\phi_{\tau T}\left(y_t^{\tau} - r_t\right)/(\tau - 1) + residual$ using daily LIBOR-Swap rates with maturities as indicated in Table 2 from August 13, 1990 to December 31, 2008. The 3-month LIBOR rate is used as the spot rate r_t.



ficients (from -0.0051 to 0.0095) than those for the latter (from -0.428 to -4.173)[4] (Dai and Singleton, 2002). It serves as a characterization of time variations in the risk premium and as a criterion for an empirically successful dynamic term structure model.

[4]This difference in the empirical patterns of violations of the "expectation hypothesis," i.e., time variations in the risk premium, raises an interesting open question that has not been studied. This issue is being investigated and results will be reported soon.

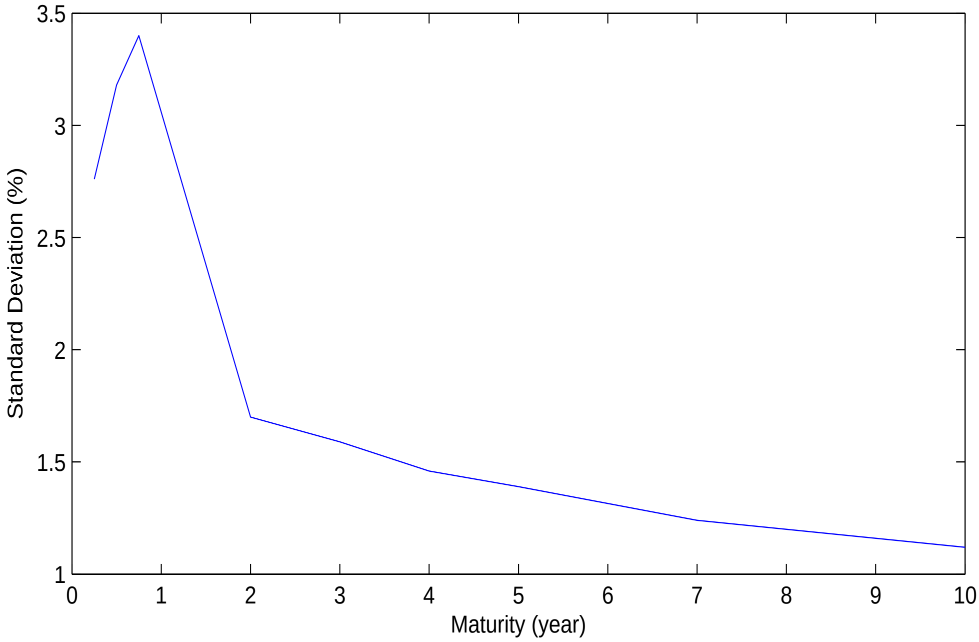
3.1.4 Time-Varying Conditional Volatility for LIBOR-Swap Yields

There exists substantial evidence that bond yields exhibit time variations in conditional second moments as well; see, for example, Ait-Sahalia (1996), Gallant and Tauchen (1998), and Andersen and Lund (1997). Hence it is also important for a dynamic term structure model to capture the time variation in the conditional volatility, which is particularly critical to the reliable valuation of many fixed-income derivatives (Dai and Singleton, 2000; Ahn et al., 2003; Ahn, Dittmar, and Gallant, 2002). According to the literature (Dai and Singleton, 2003; Buraschi, Cieslak and Trojani, 2008), there are two major issues on time variations in the conditional volatility of yields: (i), the hump-shaped term structure of yield volatilities; (ii), the degree of time variations and persistence of the conditional volatilities of yields.

Figure 3.3 , plots the historical unconditional volatility against the maturity computed as sample standard deviations of daily changes in the logarithm of the LIBOR-Swap yields. Notably, the term structure of the historical volatilities is hump-shaped with a peak around the 9-month−1-year maturity range. Furthermore, I follow Dai and Singleton (2003) to estimate a GARCH(1,1) model for the LIBOR-Swap yields. From the estimation results in Table 3.3 , it can be observed that the coefficient β, which represents the degree of volatility persistence, is fairly big, ranging from 0.6579 to 0.7960. This confirms the high degree of time variations and volatility persistence for LIBOR-Swap yields of all available maturities. Both of the two stylized facts documented above (the humped term structure and high volatility persistence) are strong evidence of time-varying conditional volatilities of yields (Dai and Singleton, 2003). Follow-

Figure 3.3: **The Term Structure of Historical Volatilities**

This figure reports the term structure of historical unconditional volatilities for daily LIBOR-Swap rates with maturities of 3-month, 6-month, 9-month, 2-year, 3-year, 4,-year, 5-year, 7-year and 10-year from August 13, 1990 to December 31, 2008. The volatility here is defined as the sample standard deviation of daily changes in the logarithm of yields.



ing Dai and Singleton (2000, 2003), Piazzesi (2005) and Buraschi, Cieslak and Trojani (2008), they will be treated as a descriptive measure of time variations in yields volatilities which an empirically successful term structure model should capture[5].

[5]Piazzesi (2005), modeling deterministic jumps by linking jump intensities directly to the meeting calendar of the Federal Open Market Committee, provides a structural (monetary) interpretation of the volatility hump.

Table 3.3: **GARCH(1,1) Parameters for the LIBOR-Swap Yields**

This table reports the maximum likelihood estimates of a GARCH(1,1) model: $\sigma_t^2 = \overline{\sigma} + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2$, where ε_t is the innovation from the AR(1) representation of the LIBOR-Swap yields with maturities as indicated in the table from August 13, 1990 to December 31, 2008. Standard errors are given in the "SE" columns.

GARCH(1,1)	$\overline{\sigma}$		α		β	
Maturity	Est.	SE	Est.	SE	Est.	SE
3-month	0.0001	$3.8 \cdot 10^{-5}$	0.2486	0.0637	0.7523	0.0639
6-month	0.0002	$3.9 \cdot 10^{-5}$	0.2112	0.1282	0.7851	0.1180
9-month	0.0002	$3.4 \cdot 10^{-5}$	0.1989	0.0937	0.7960	0.0989
2-year	0.0003	$6.5 \cdot 10^{-5}$	0.1723	0.1059	0.7902	0.1009
3-year	0.0001	$1.2 \cdot 10^{-5}$	0.1706	0.1139	0.7990	0.1095
4-year	0.0002	$1.9 \cdot 10^{-5}$	0.1350	0.0761	0.7419	0.0653
5-year	0.0001	$2.6 \cdot 10^{-5}$	0.1679	0.0134	0.7780	0.1239
7-year	0.0005	$5.1 \cdot 10^{-5}$	0.1464	0.0771	0.6579	0.0592
10-year	0.0002	$4.7 \cdot 10^{-5}$	0.1564	0.0169	0.7805	0.1555

3.2 AJD Term Structure Models

3.2.1 General Model Specifications

The specification of a general multifactor AJD term structure model, consisting of risk-neutral dynamics, market prices of risks, and physical dynamics, is described in this section. Such a multifactor AD term structure model is a special case of the AJD model without the jumps specifications. A multivariate AJD model specifies the instantaneous risk-free rate r_t as

$$r_t = \delta_0 + \delta_1' X_t \tag{3.4}$$

which is a deterministic affine function of the $d \times 1$ vector of state variables X_t. Here δ_0 is a scalar and δ_1 is an $d \times 1$ vector in $\mathbb{R}^d$. Under an equivalent martingale measure Q, also called risk-neutral measure, the multivariate state variable $\{X_t\}$ is a Markov process defined by the following stochastic differential eqnarray (SDE):

$$dX_t = \widetilde{\mathcal{K}}\left(\widetilde{\theta} - X_t\right) dt + \Sigma \sqrt{S_t} dW_t^Q + dJ_t^Q, \tag{3.5}$$

where $\widetilde{\theta}$ is a $d \times 1$ vector, $\widetilde{\mathcal{K}}$ and Σ are $d \times d$ matrices, which may be non-diagonal and asymmetric, S_t is a diagonal matrix with the i-th diagonal element given by

$$[S_t]_{ii} = \alpha_i + \beta_i' X_t, \tag{3.6}$$

W_t^Q is the $d \times 1$ standard Brownian motion, and J_t^Q is a pure jump process with the jump arrival intensity

$$\lambda^Q(X_t) = \lambda_0^Q + \left(\lambda_1^Q\right)' X_t \tag{3.7}$$

for a scalar λ_0^Q and a $d \times 1$ vector λ_1^Q and with a d-dimensional random jump size vector ξ^Q having the mean μ^Q. The term "affine" refers to the fact that the instantaneous short rate in (3.4), the drift term in (3.5), the conditional variance in (3.6) and the jump arrival intensity in (3.7) are all affine functions[6] of the state variables X_t.

[6]Such affine specifications do not constrain the short rate and jump intensities to be positive in general since the state variables X_t may take negative values. See Piazzesi (2010, Section 3.6) for detailed discussions about negative short rates and jump intensities.

Compared with the AD term structure models, defined in the same way as the AJD models above except with state variables following

$$dX_t = \widetilde{\mathcal{K}}\left(\widetilde{\theta} - X_t\right)dt + \Sigma\sqrt{S_t}dW_t^Q,$$

which do not incorporate jumps into the state variable dynamics, the AJD models capture the potential jump activities via the pure-jump process J_t^Q with two components: random jump-occurrence times and random jump sizes. Suppose the jump-occurrence times are $\{\Gamma_i, i \geq 1\}$. Then in the specified dynamics (3.5) for the state variables, jumps happen with a state-dependent stochastic intensity process $\left\{\lambda^Q(X_t)\right\}$.[7] Given the occurrence of the i-th jump, the state variable X_t jumps from $X(\Gamma_i-)$ to $X(\Gamma_i-)\,\xi_i^Q$, where ξ_i^Q is independent of W_t^Q and ξ_j^Q for $i \neq j$. The intuition is that the conditional probability at time t of another jump before time $t+\Delta$ is approximately $\lambda^Q\Delta$ for small Δ and, given a jump-occurrence, the mean relative jump size is $\mu^Q = E\left[\xi^Q - 1\right]$. Hence, the last term $\left(\lambda^Q\mu^Q\right)dt$ in (3.5) combines the effects of random jump timing and sizes and acts as the compensator for the instantaneous change in X_t induced by the pure jump process J_t^Q.

The absence of arbitrage opportunities implies that the time-t price of a zero-coupon bond which matures at time T is given by

$$P(X_t, t, T) = E^Q\left[\exp\left(-\int_t^T r_u du\right)\middle| X_t\right], \tag{3.8}$$

where the expectation is taken under the risk-neutral dynamics of X_t defined in (3.5). Following from Duffie, Pan, and Singleton (2000), the bond price P satisfies the following partial differential eqnarray:

$$r_t P - \frac{\partial P}{\partial t} = \left[\widetilde{\mathcal{K}}\left(\widetilde{\theta} - X_t\right) - \lambda^Q\mu^Q\right]' \left.\frac{\partial P}{\partial x}\right|_{x=X_t}$$

[7] Such a specification of the jump activity is of the Cox-process type. See Brémaud(1981).

$$+\frac{1}{2}Trace\left[\Omega_t \left.\frac{\partial^2 P}{\partial x \partial x'}\right|_{x=X_t}\right] + \lambda^Q E\left[P\left(X_t + \xi^Q, t, T\right) - P\left(X_t, t, T\right)\right], \quad (3.9)$$

where $\Omega_t = \Sigma S_t \Sigma'$. Eqnarray (3.9) is actually derived by Ito's Lemma and $E_t^Q\left[\frac{dP}{P}\right] = r_t$; see Duffie, Pan, and Singleton (2000) and Shreve (2004) for technical details of Ito's Lemma for jump processes and the derivation of (3.9). Of course, the bond price has to satisfy a final condition: $P(X_t, T, T) = 1$ for all T. This model is well-defined as long as the risk-neutral dynamics of X_t in (3.5) are well-defined and the expected value in (3.8) is finite. Equivalently, this implies that the eqnarray (3.9) has a well-defined solution under some additional technical regularity conditions.

Duffie, Pan, and Singleton (2000) show that the bond prices actually assume the exponential affine form:

$$P(X_t, t, \tau) = \exp\left[A(\tau) - B(\tau)' X_t\right] \quad (3.10)$$

where $A(\tau)$ and $B(\tau) = [B_1(\tau), \cdots, B_d(\tau)]'$ satisfy the so-called complex-valued Riccati-type ordinary differential eqnarrays:

$$\begin{aligned}
\frac{dA(\tau)}{d\tau} &= -\left(\widetilde{\mathcal{K}}\widetilde{\theta}\right)' B(\tau) + \frac{1}{2}\sum_{i=1}^{N}\left[\Sigma' B(\tau)\right]_i^2 \alpha_i - \delta_0 + \lambda_0^Q\left[\varsigma^Q(B(\tau)) - 1\right] \\
\frac{dB(\tau)}{d\tau} &= -\widetilde{\mathcal{K}}' B(\tau) - \frac{1}{2}\sum_{i=1}^{N}\left[\Sigma' B(\tau)\right]_i^2 \beta_i + \delta_1 + \lambda_1^Q\left[\varsigma^Q(B(\tau)) - 1\right] \quad (3.11)
\end{aligned}$$

where $\tau = T - t$ is the bond's time to maturity and $\varsigma^Q(u) = E\left[\exp\left(u'\xi^Q\right)\right]$ is the moment-generating function of the random jump size vector ξ^Q. These ordinary differential equations can be solved easily through numerical integration techniques, starting from the initial conditions $A(0) = 0$ and $B(0) = 0_{d\times 1}$. By (3.10), the yields of zero coupon bonds, $y(X_t, \tau) \equiv -\frac{1}{\tau}\log\left[P(X_t, t, \tau)\right]$, are also affine functions of the state variables[8]:

$$y(X_t, \tau) = \frac{1}{\tau}\left[-A(\tau) + B(\tau)' X_t\right] \tag{3.12}$$

To employ the closed-form formulas of eqnarray (3.10) in empirical studies of the AJD term structure models, we also need to know the dynamics of X_t and $P(X_t, t, \tau)$ under the physical probability measure $\mathcal{P}$. This further requires us to specify the market prices of risk due to both the diffusive risk (the volatility uncertainty) and the jump risk. For the market price of diffusive risks Λ_t, there are usually two types of specifications. The first, employed by Dai and Singleton (2000) and characterized as "completely" affine specifications, assumes that

$$\Lambda_t = \sqrt{S_t}\eta_1 \tag{3.13}$$

where η_1 is an $d \times 1$ parameter vector. This implies that the compensation for the diffusion risk is a fixed multiple of the variance risk. This restriction makes it difficult to replicate some stylized facts about historical excess bond returns (Dai and Singleton, 2002, 2003; Duffee, 2002). As a result, completely affine models may provide poor forecasts of future bond yields and forecast errors may be large, especially when the slope of the term structure is steep. Duffee (2002) proposes the "essentially" affine specifications by assuming

$$\Lambda_t = \sqrt{S_t}\eta_1 + \sqrt{S_t^-}\eta_2 X_t \tag{3.14}$$

where S_t^- is an $d \times d$ diagonal matrix with the (i, i)-th element

$$[S_t^-]_{ii} = \begin{cases} \left(\alpha_i + \beta_i' X_t\right)^{-1}, if \inf\left(\alpha_i + \eta_i' X_t\right) > 0 \\ 0, \qquad otherwise \end{cases} \quad i = 1 \cdots, d \tag{3.15}$$

[8]Alternatively, one can start with the requirement that the yields be affine and show that the dynamics of the state vector X_t must be affine (see Duffie and Kan (1996) for similar derivations in AD models without jumps).

and η_2 is an $d \times d$ matrix. It can be seen from (3.15) that Λ_t can depend on X_t directly and the sign for the market price of risk can change now, representing an obvious improvement over (3.13).

For the market prices of jump risks, I follow Pan (2002) and Jarrow, Li, and Zhao (2006) to first specify the dynamics of X_t under the physical measure and then take the difference in the jump dynamics under risk-neutral and physical measures as the jump risk premium. Under both specifications of Λ_t in equations (3.16) and (3.17), we assume that X_t still follows an AJD model under the physical measure:

$$dX_t = \widetilde{\mathcal{K}}\left(\widetilde{\theta} - X_t\right)dt + \Sigma\sqrt{S_t}dW_t + \Sigma\sqrt{S_t}\Lambda_t dt + dJ_t, \tag{3.16}$$

where W_t is the $d \times 1$ standard Brownian Motion under the physical measure $\mathcal{P}$, J_t is a pure jump process with the jump arrival intensity

$$\lambda(X_t) = \lambda_0 + (\lambda_1)' X_t \tag{3.17}$$

for a scalar λ and a $d \times 1$ vector λ_1 and with a d-dimensional random jump size vector ξ having the mean μ, and $(\lambda\mu)\,dt$ is the compensator for the pure jump process J_t under the physical measure. It can be seen that the third term in (3.16), $\Sigma\sqrt{S_t}\Lambda_t dt$, represents the market price of diffusive risk.

Now we can discuss the market price of jump risks. Comparing the physical dynamics of X_t in (3.16) and the risk-neutral dynamics in (3.5), we see that by allowing the risk-neutral mean relative jump size $\mu^{\mathcal{Q}}$ to differ from its physical counterpart μ, a premium is assigned for the jump-size uncertainty. Similarly, a premium is also accommodated for the jump-timing risk by permitting the coefficients $\left(\lambda_0^{\mathcal{Q}}, \lambda_1^{\mathcal{Q}}\right)$ in the risk-neutral jump-arrival intensity to differ from their physical counterparts (λ_0, λ_1).[9] Therefore, the time-t compensation for the jump-risk is

$$
\begin{aligned}
& [\lambda_0 + (\lambda_1)' X_t]\,\mu - \left[\lambda_0^Q + \left(\lambda_1^Q\right)' X_t\right]\mu^Q \\
= \; & \lambda_0\mu - \lambda_0^Q\mu^Q + \left[(\lambda_1)'\mu - \left(\lambda_1^Q\right)'\mu^Q\right] X_t
\end{aligned} \tag{3.18}
$$

In contrast, Piazzesi (2005) set the market prices of jump timing uncertainty for target-rate moves to zero considering the data limitation (only 5 years of data). Overall, combining (3.14) and (3.18) yields the total risk premium in a multifactor AJD term structure model:

$$
\Lambda_t^{AJD} = \sqrt{S_t}\eta_1 + \sqrt{S_t^-}\eta_2 X_t + \lambda_0\mu - \lambda_0^Q\mu^Q + \left[(\lambda_1)'\mu - \left(\lambda_1^Q\right)'\mu^Q\right] X_t, \tag{3.19}
$$

where $\lambda_0\mu - \lambda_0^Q\mu^Q$ is a constant, $\sqrt{S_t}\eta_1$ depends only on $\sqrt{S_t}$ and $\left[\sqrt{S_t^-}\eta_2 + (\lambda_1)'\mu - \left(\lambda_1^Q\right)'\mu^Q\right] X_t$ depends on X_t directly.

3.2.2 Theoretical Time Variations in the Risk Premium and Conditional Volatility

In this section, I shall provide initial clues about what causes the empirical performance of the AJD and AD models to vary when capturing time variations in the risk premium and conditional volatility. In particular, the theoretical time variability of both the risk premium and conditional volatility will be analyzed. First, for the AJD term structure model in the previous section, the instantaneous expected bond return (see Piazzesi (2010) for the derivation) is

$$
\mu^e(t,\tau) = -B(\tau)'\Sigma\sqrt{S_t}\left[\sqrt{S_t}\eta_1 + \sqrt{S_t^-}\eta_2 X_t\right] + \lambda(X_t)\left[\varsigma^Q(B(\tau)) - 1\right],
$$

[9] Pan (2002), in studying the jump-risk premiums using the S&P 500 index and near-the-money short-dated option prices, specifies a premium only for the jump size uncertainty. In that case, the jump risk premiums are all artificially absorbed by the jump size risk premium coefficient $\mu - \mu^Q$.

where the first term is related to the market prices of diffusive risk and the second is related to the market prices of jump risk. This implies that we need a flexible specification of the market prices of risk to capture the time variation in the risk premium or the expected excess bond returns.

For AD models, the market price of risk is

$$\Lambda_t^{AD} = \sqrt{S_t}\eta_1 + \sqrt{S_t^-}\eta_2 X_t$$

by (3.19). The first term $\sqrt{S_t}\eta_1$, which is proposed by Dai and Singleton (2000) in the so-called "completely" affine models, is very restrictive in that the sign of the i-th element is the same as the sign of the i-th element of the vector η_1. Therefore, the sign of any element in the first term of Λ_t^{AD} cannot change over time. The second term $\sqrt{S_t^-}\eta_2 X_t$ proposed by Duffee (2002) in the "essentially" affine models partially solves this limitation: The i-th element of $\sqrt{S_t^-}\eta_2 X_t$ can change sign if $[S_t^-]_{ii} \neq 0$ since it will depend on the i-th element of the state factor X_t which is able to switch signs itself. Hence some elements of the market price of diffusive risk Λ_t^{AD}can change signs over time.

There is, however, a restriction on the specification of S_t^- in the form of (3.15), which shows that S_t^- is closely linked to the conditional volatility specification S_t. In fact, the restriction reveals that the more factors there are in the conditional volatility S_t, the more zeros there are in the elements of S_t^- and the fewer elements there are in Λ_t^{AD}, which can change sign over time. This actually explains the findings in Dai and Singleton (2002) and Duffee (2002) that AD models cannot simultaneously capture time variations in the risk premium and conditional volatility for U.S. Treasury yields.

In contrast, the market price of risk in AJD models Λ_t^{AJD} has an additional term $\lambda_0\mu - \lambda_0^Q\mu^Q + \left[(\lambda_1)'\mu - \left(\lambda_1^Q\right)'\mu^Q\right]X_t$ that contains compensations for both the jump size and the jump timing risks. It can be observed that the term $\left[(\lambda_1)'\mu - \left(\lambda_1^Q\right)'\mu^Q\right] X_t$ depends on all elements of the state vector X_t directly, and is able to switch sign over time in a very flexible way. More importantly, this term is introduced only by the jumps and has nothing to do with the conditional volatility S_t. Hence, jump risk premiums generalize the market prices of risks significantly without imposing a single restriction on the conditional volatility. This is exactly the key that allows AJD term structure models to simultaneously match time variations in both the risk premium and conditional volatility.

3.2.3 Three-Factor Models

For general AJD term structure models in Section 3.2.1, the parameters $\widetilde{\mathcal{K}}$, $\widetilde{\theta}$, Σ, α, and β cannot be chosen arbitrarily. Based on Dai and Singleton (2000), we assume that some admissibility conditions are required for the existence of the process X_t. They actually show that there exist $d+1$ disjoint admissible regions of the parameter space for each d. Denote β the $d \times d$ matrix with the i-th column as the vector β_i in (3.6) and M as the rank of β, which is the number of independent linear combinations of state variables entering the conditional volatility specification. Then with d factors, there are $d+1$ non-nested families of AJD models corresponding to $M = 0, 1, \cdots d$ and denoted as $\text{AJD}_M(\text{d})$, which impose a range of restrictions on the parameters $\widetilde{\mathcal{K}}$, $\widetilde{\theta}$, Σ, α, and β. Since there exist several specifications of the model parameters that generate identical dynamics of interest rates, I follow Dai and Singleton (2000) in considering the canonical representation for each family of AJD models with the matrix Σ normalized as the identity matrix. This normalization does not result in any loss of generality because, for a Σ that is distinct from the identity matrix, we can construct a new set of state variables $Z_t = \Sigma^{-1}X_t$, whose volatility matrix is then diagonal.

The canonical representations of the d-dimensional AJD term structure models will be presented in the following under the essentially affine market prices of diffusion risk (Duffee, 2002) in (3.14), leading to "essentially" AJD term structure models. Given the risk-neutral dynamics of the state variables in (3.5), the physical dynamics of the state variables for essentially affine specifications are, by (3.14)-(3.16),

$$\begin{aligned} dX_t &= \widetilde{\mathcal{K}}\left(\widetilde{\theta} - X_t\right)dt + \Sigma\sqrt{S_t}dW_t^P + \Sigma S_t\eta_1 dt + \Sigma\sqrt{S_t}\sqrt{S_t^-}\eta_2 X_t dt + dJ_t \\ &= \mathcal{K}_E\left(\theta_E - X_t\right)dt + \Sigma\sqrt{S_t}dW_t^P + dJ_t, \end{aligned}$$

where

$$\mathcal{K}_E = \widetilde{\mathcal{K}} - \Sigma\begin{pmatrix} \eta_{1,1}\eta_1' \\ \vdots \\ \eta_{1,N}\eta_N' \end{pmatrix} + S_t S_t^- \eta_2, \theta_E = \mathcal{K}^{-1}\left(\widetilde{\mathcal{K}}\widetilde{\theta} + \Sigma\begin{pmatrix} \eta_{1,1}\alpha_1 \\ \vdots \\ \eta_{1,N}\alpha_N \end{pmatrix}\right).$$

Then for each M, we partition X_t as $X_t = \left(X_t^{B\prime}, X_t^{D\prime}\right)'$ where X_t^B and X_t^D are $M\times 1$- and $(d-M)\times 1$-dimensional, respectively. The corresponding representations for the parameters are

$$\mathcal{K}_E = \begin{bmatrix} \mathcal{K}_{M\times M}^{BB} & 0_{M\times(d-M)} \\ \mathcal{K}_{(d-M)\times M}^{DB} & \mathcal{K}_{(d-M)\times(d-M)}^{DD} \end{bmatrix} for M > 0$$

and either upper or lower triangular for $M = 0$,

$$\begin{aligned} \theta_E &= \begin{pmatrix} \theta_{M\times 1}^B \\ 0_{(d-M)\times 1} \end{pmatrix}, \alpha = \begin{pmatrix} 0_{M\times 1} \\ 1_{(d-M)\times 1} \end{pmatrix}, \\ B &= (\beta_1, ..., \beta_d) = \begin{bmatrix} I_{M\times M} & B_{M\times(d-M)}^{BD} \\ 0_{(d-M)\times M} & 0_{(d-M)\times(d-M)} \end{bmatrix} \end{aligned}$$

and $\Sigma = I$. Given these identification normalization conditions, constraints on the parameters for the existence of the process and non-attainment of the boundaries can be obtained (Dai and Singleton, 2000; Ait-Sahalia and Kimmel, 2010), which deliver the the canonical representation of d-dimensional AJD term structure models. We classify all admissible three-factor essentially AJD (denoted as $\text{EAJD}_M(3)$) term structure models into subfamilies and within each subfamily present the maximal model in the following. They are in fact three-factor essentially AD models in Duffee (2002) augmented by jumps. Note that for all subfamilies, the sport interest rate is always specified as $r_t = \delta_0 + \delta_{11}X_{1t} + \delta_{12}X_{2t} + \delta_{13}X_{3t}$.

- $\text{EAJD}_0(3)$ Model: In this model, the dynamics of X_t under the physical measure are given as

$$d\begin{bmatrix} X_{1t} \\ X_{2t} \\ X_{3t} \end{bmatrix} = \begin{bmatrix} \kappa_{11} & 0 & 0 \\ \kappa_{21} & \kappa_{22} & 0 \\ \kappa_{31} & \kappa_{32} & \kappa_{33} \end{bmatrix} \begin{bmatrix} -X_{1t} \\ -X_{2t} \\ -X_{3t} \end{bmatrix} dt + d\begin{bmatrix} W_{1t}^P \\ W_{2t}^P \\ W_{3t}^P \end{bmatrix} + dJ_t, \tag{3.20}$$

  where J_t is a 3-dimensional pure jump process, with the jump arrival intensity $\lambda(X_t) = \lambda_0 + \lambda_{11}X_{1t} + \lambda_{12}X_{2t} + \lambda_{13}X_{3t}$ and the random jump size vector ξ having the mean vector $\mu = (\mu_1, \mu_2, \mu_3)'$.

Since $M = 0$, none of the elements in X_t affect the volatility of X_t and hence the state variables are homoscedastic. The corresponding risk neutral dynamics for $\text{EAJD}_0(3)$ are

$$d\begin{bmatrix} X_{1t} \\ X_{2t} \\ X_{3t} \end{bmatrix} = \begin{bmatrix} \kappa_{11} & 0 & 0 \\ \kappa_{21} & \kappa_{22} & 0 \\ \kappa_{31} & \kappa_{32} & \kappa_{33} \end{bmatrix} \begin{bmatrix} -X_{1t} \\ -X_{2t} \\ -X_{3t} \end{bmatrix} dt - \begin{bmatrix} \eta_{11} \\ \eta_{12} \\ \eta_{13} \end{bmatrix} d$$

$$-\begin{bmatrix} \eta_{2,11} & \eta_{2,12} & \eta_{2,13} \\ \eta_{2,21} & \eta_{2,22} & \eta_{2,23} \\ \eta_{2,31} & \eta_{2,32} & \eta_{2,33} \end{bmatrix} \begin{bmatrix} X_{1t} \\ X_{2t} \\ X_{3t} \end{bmatrix} dt + d \begin{bmatrix} W_{1t}^P \\ W_{2t}^P \\ W_{3t}^P \end{bmatrix} + dJ_t^Q \qquad (3.21)$$

where J_t^Q is a 3-dimensional pure jump process, with the jump arrival intensity $\lambda^Q(X_t) = \lambda_0^Q + \lambda_{11}^Q X_{1t} + \lambda_{12}^Q X_{2t} + \lambda_{13}^Q X_{3t}$ and the random jump size vector ξ^Q having the mean vector $\mu = \left(\mu_1^Q, \mu_2^Q, \mu_3^Q\right)'$. If the terms related to jumps in (3.20)-(3.21) are taken off, we get the physical and risk-neutral dynamics of the $EA_0(3)$ model in the three-factor essentially AD class of Dai and Singleton (2002) and Duffee (2002).

- $EAJD_1(3)$ Model: In this model, the dynamics of X_t under the physical measure are given as

$$d\begin{bmatrix} X_{1t} \\ X_{2t} \\ X_{3t} \end{bmatrix} = \begin{bmatrix} \kappa_{11} & 0 & 0 \\ \kappa_{21} & \kappa_{22} & \kappa_{23} \\ \kappa_{31} & \kappa_{32} & \kappa_{33} \end{bmatrix} \begin{bmatrix} \theta_1 - X_{1t} \\ -X_{2t} \\ -X_{3t} \end{bmatrix} dt$$
$$+ \begin{bmatrix} \sqrt{X_{1t}} & & \\ & \sqrt{1+\beta_{21}X_{1t}} & \\ & & \sqrt{1+\beta_{31}X_{1t}} \end{bmatrix} \begin{bmatrix} dW_{1t}^P \\ dW_{2t}^P \\ dW_{3t}^P \end{bmatrix} + dJ_t, \qquad (3.22)$$

where J_t is a 3-dimensional pure jump process, with the jump arrival intensity $\lambda(X_t) = \lambda_0 + \lambda_{11}X_{1t} + \lambda_{12}X_{2t} + \lambda_{13}X_{3t}$ and the random jump size vector ξ having the mean vector $\mu = (\mu_1, \mu_2, \mu_3)'$.

Since $M = 1$, the first element of the state variable X_t, X_{1t}, determines the conditional volatility of all three state variables. The corresponding risk-neutral dynamics for $\text{EAJD}_1(3)$ are

$$d\begin{bmatrix} X_{1t} \\ X_{2t} \\ X_{3t} \end{bmatrix} = \begin{bmatrix} \kappa_{11} & 0 & 0 \\ \kappa_{21} & \kappa_{22} & \kappa_{23} \\ \kappa_{31} & \kappa_{32} & \kappa_{33} \end{bmatrix}\begin{bmatrix} \theta_1 - X_{1t} \\ -X_{2t} \\ -X_{3t} \end{bmatrix} dt - \begin{bmatrix} X_{1t}\eta_{11} \\ (1+\beta_{21}X_{1t})\,\eta_{12} \\ (1+\beta_{31}X_{1t})\,\eta_{13} \end{bmatrix} dt$$
$$- \begin{bmatrix} 0 & 0 & 0 \\ \eta_{2,21} & \eta_{2,22} & \eta_{2,23} \\ \eta_{2,31} & \eta_{2,32} & \eta_{2,33} \end{bmatrix}\begin{bmatrix} X_{1t} \\ X_{2t} \\ X_{3t} \end{bmatrix} dt$$
$$+ \begin{bmatrix} \sqrt{X_{1t}} & & \\ & \sqrt{1+\beta_{21}X_{1t}} & \\ & & \sqrt{1+\beta_{31}X_{1t}} \end{bmatrix}\begin{bmatrix} dW^Q_{1t} \\ dW^Q_{2t} \\ dW^Q_{3t} \end{bmatrix} + dJ^Q_t, \tag{3.23}$$

where J^Q_t is a 3-dimensional pure jump process, with the jump arrival intensity $\lambda^Q(X_t) = \lambda^Q_0 + \lambda^Q_{11}X_{1t} + \lambda^Q_{12}X_{2t} + \lambda^Q_{13}X_{3t}$ and the random jump size vector ξ^Q having the mean vector $\mu = \left(\mu^Q_1, \mu^Q_2, \mu^Q_3\right)'$. If the terms related to jumps in both (3.22) and (3.23) are taken off, we get the physical and risk-neutral dynamics of the $\text{EA}_1(3)$ model in the three-factor essentially AD class of Dai and Singleton (2002) and Duffee (2002).

- $\text{EAJD}_2(3)$ Model: In this model, the dynamics of X_t under the physical measure are given as

$$d\begin{bmatrix} X_{1t} \\ X_{2t} \\ X_{3t} \end{bmatrix} = \begin{bmatrix} \kappa_{11} & \kappa_{12} & 0 \\ \kappa_{21} & \kappa_{22} & 0 \\ \kappa_{31} & \kappa_{32} & \kappa_{33} \end{bmatrix}\begin{bmatrix} \theta_1 - X_{1t} \\ \theta_2 - X_{2t} \\ -X_{3t} \end{bmatrix} dt$$
$$+ \begin{bmatrix} \sqrt{X_{1t}} & & \\ & \sqrt{X_{2t}} & \\ & & \sqrt{1+\beta_{31}X_{1t}+\beta_{32}X_{2t}} \end{bmatrix}\begin{bmatrix} dW^P_{1t} \\ dW^P_{2t} \\ dW^P_{3t} \end{bmatrix} + dJ_t \tag{3.24}$$

where J_t is a 3-dimensional pure jump process, with the jump arrival intensity $\lambda(X_t) = \lambda_0 + \lambda_{11}X_{1t} + \lambda_{12}X_{2t} + \lambda_{13}X_{3t}$ and the random jump size vector ξ having the mean vector $\mu = (\mu_1, \mu_2, \mu_3)'$.

Since $M = 2$, the first two elements of the state variable X_t, X_{1t} and X_{2t}, determine the conditional volatility of all three state variables. The corresponding risk-neutral dynamics for $\mathrm{EAJD}_2(3)$ are

$$d\begin{bmatrix} X_{1t} \\ X_{2t} \\ X_{3t} \end{bmatrix} = \begin{bmatrix} \kappa_{11} & \kappa_{12} & 0 \\ \kappa_{21} & \kappa_{22} & 0 \\ \kappa_{31} & \kappa_{32} & \kappa_{33} \end{bmatrix}\begin{bmatrix} \theta_1 - X_{1t} \\ \theta_2 - X_{2t} \\ -X_{3t} \end{bmatrix} dt - \begin{bmatrix} X_{1t}\eta_{11} \\ X_{2t}\eta_{12} \\ (1 + \beta_{31}X_{1t} + \beta_{32}X_{2t})\,\eta_{13} \end{bmatrix} dt$$
$$- \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ \eta_{2,31} & \eta_{2,32} & \eta_{2,33} \end{bmatrix}\begin{bmatrix} X_{1t} \\ X_{2t} \\ X_{3t} \end{bmatrix} dt$$
$$+ \begin{bmatrix} \sqrt{X_{1t}} & & \\ & \sqrt{X_{2t}} & \\ & & \sqrt{1 + \beta_{31}X_{1t} + \beta_{32}X_{2t}} \end{bmatrix}\begin{bmatrix} dW^Q_{1t} \\ dW^Q_{2t} \\ dW^Q_{3t} \end{bmatrix} + dJ^Q_t \tag{3.25}$$

where J^Q_t is a 3-dimensional pure jump process, with the jump arrival intensity $\lambda^Q(X_t) = \lambda^Q_0 + \lambda^Q_{11}X_{1t} + \lambda^Q_{12}X_{2t} + \lambda^Q_{13}X_{3t}$ and the random jump size vector ξ^Q having the mean vector $\mu = \left(\mu^Q_1, \mu^Q_2, \mu^Q_3\right)'$. If the terms related to jumps in both (3.24) and (3.25) are taken off, we get the physical and risk-neutral dynamics of the $\mathrm{EA}_2(3)$ model in the three-factor essentially AD class of Dai and Singleton (2002) and Duffee (2002).

- $\mathrm{AJD}_3(3)$ Model: In this model, the dynamics of X_t under the physical measure are given as

$$d\begin{bmatrix} X_{1t} \\ X_{2t} \\ X_{3t} \end{bmatrix} = \begin{bmatrix} \kappa_{11} & \kappa_{12} & \kappa_{13} \\ \kappa_{21} & \kappa_{22} & \kappa_{23} \\ \kappa_{31} & \kappa_{32} & \kappa_{33} \end{bmatrix} \begin{bmatrix} \theta_1 - X_{1t} \\ \theta_2 - X_{2t} \\ \theta_3 - X_{3t} \end{bmatrix} dt + \begin{bmatrix} \sqrt{X_{1t}} & & \\ & \sqrt{X_{2t}} & \\ & & \sqrt{X_{3t}} \end{bmatrix} \begin{bmatrix} dW_{1t}^P \\ dW_{2t}^P \\ dW_{3t}^P \end{bmatrix} + dJ_t, \tag{3.26}$$

where J_t is a 3-dimensional pure jump process, with the jump arrival intensity $\lambda(X_t) = \lambda_0 + \lambda_{11}X_{1t} + \lambda_{12}X_{2t} + \lambda_{13}X_{3t}$ and the random jump size vector ξ having the mean vector $\mu = (\mu_1, \mu_2, \mu_3)'$.

Since $M = 3$, all three components of the state variable X_t enter the conditional volatility of the state variables. The risk-neutral dynamics for $\text{EAJD}_3(3)$ are

$$d\begin{bmatrix} X_{1t} \\ X_{2t} \\ X_{3t} \end{bmatrix} = \begin{bmatrix} \kappa_{11} & \kappa_{12} & \kappa_{13} \\ \kappa_{21} & \kappa_{22} & \kappa_{23} \\ \kappa_{31} & \kappa_{32} & \kappa_{33} \end{bmatrix} \begin{bmatrix} \theta_1 - X_{1t} \\ \theta_2 - X_{2t} \\ \theta_3 - X_{3t} \end{bmatrix} dt - \begin{bmatrix} X_{1t}\eta_{11} \\ X_{2t}\eta_{12} \\ X_{3t}\eta_{13} \end{bmatrix} dt + \begin{bmatrix} \sqrt{X_{1t}} & & \\ & \sqrt{X_{2t}} & \\ & & \sqrt{X_{3t}} \end{bmatrix} \begin{bmatrix} dW_{1t}^Q \\ dW_{2t}^Q \\ dW_{3t}^Q \end{bmatrix} + dJ_t^Q, \tag{3.27}$$

where J_t^Q is a 3-dimensional pure jump process, with the jump arrival intensity $\lambda^Q(X_t) = \lambda_0^Q + \lambda_{11}^Q X_{1t} + \lambda_{12}^Q X_{2t} + \lambda_{13}^Q X_{3t}$ and the random jump size vector ξ^Q having the mean vector $\mu = \left(\mu_1^Q, \mu_2^Q, \mu_3^Q\right)'$. If the terms related to jumps in both (3.26) and (3.27) are taken off, we get the physical and risk-neutral dynamics of the $\text{EA}_3(3)$ model in the three-factor essentially AD class of Dai and Singleton (2002) and Duffee (2002).

3.3 Infinitesimal Operator Methods for Model Estimation

The description of the estimation method is divided into two sections. First, the infinitesimal operator method proposed in Chapter 2 for a general multivariate continuous time Markov model without latent state variables are introduced. Second, this method is adapted to estimate the affine term structure models with unobservable risk factors. This infinitesimal operator based method is equivalent to MLE in the sense of employing the same information set about the process dynamics but is more convenient numerically and computationally due to its closed-form expression which the transition density does not have.

3.3.1 Models without Latent Factors

Consider a multivariate time-homogeneous Markov model defined by the following stochastic differential eqnarray on the filtered probability space $(\Omega, \mathfrak{F}, \{\mathfrak{F}_t\}_{t\geq 0}, \mathcal{P})$:

$$dX_t = b(X_t;\theta)dt + \sigma(X_t;\theta)dW_t + dJ_t \tag{3.28}$$

where W_t is a $d \times 1$ standard Brownian motion in $\mathbb{R}^d$, $b : E \subset \mathbb{R}^d \to \mathbb{R}^d$ is a drift function (i.e., the instantaneous conditional mean), $\sigma : E \to \mathbb{R}^{d\times d}$ is a volatility function (i.e., the instantaneous conditional standard deviation), and $\Theta \subset \mathbb{R}^p$ is a finite-dimensional parameter space. The jump process J_t is of a Poisson-type with the jump arrival intensity $\lambda(X_t, \theta)$ and the random jump size vector ξ_t, which is independent of $\mathfrak{F}_{t-}$ and has probability density $\nu(\cdot, \theta) : \mathbb{R}^d \to \mathbb{R}$.[10]

[10]The framework in Chapter 2 actually allows J_t to be an infinite-activity pure jump Levy process, which accommodates an infinite number of jumps within any finite time interval.

The model in (3.28) is a full parametric model in the sense that the transition density is specified. It is natural that maximum likelihood methods are the preferred econometric tools due to their advantageous statistical properties, such as efficiency. However, as discussed earlier, the transition density of most continuous-time Markov models have no analytic expressions and this raises serious obstacles to the implementation. It is more convenient to employ an alternative characterization tool that is available in closed form, i.e., the infinitesimal operator defined as follows for the model in (3.28) (Rogers and Williams, 2000):

$$\begin{aligned} \mathcal{A}f(x) &= \sum_{i=1}^{d} b_i(x;\theta) f_i'(x) + \frac{1}{2}\sum_{i,j=1}^{d} a_{i,j}(x;\theta) f_{i,j}''(x) \\ &\quad + \lambda(x,\theta) \int \left[f(x+c) - f(x) \right] d\nu(c,\theta), \end{aligned} \tag{3.29}$$

where $x \in \mathbb{R}^d$, $f \in D(\mathcal{A})$ which is the domain of $\mathcal{A}$, and

$$a_{ij}(x;\theta) = \sum_{k=1}^{d} \sigma_{i,k}(x;\theta)\sigma_{j,k}(x;\theta). \tag{3.30}$$

It is obvious from (3.29)-(3.30) that the infinitesimal operator is always analytic. Furthermore, it fully characterizes the dynamics of the model and is equivalent to the transition density in this sense.[11]

To utilize the infinitesimal operator for econometric inferences, a transformation is considered based on the celebrated "martingale problem," which is defined as follows (Karatzas and Shreve, 1991): A probability measure $\mathcal{P}$ under which

Therefore, the model framework in (3.28) is actually general enough to cover most popular continuous-time financial models, such as diffusion, jump diffusion, and Levy-type jump diffusion models, in option pricing, term structures of interest rates, and exchange rate dynamics. See Sundaresan (2000) for a general survey of continuous-time finance, Dai and Singleton (2003) for term structure models, and Wu (2008) for Levy-type models. I focus on Poisson-type jumps here since these are the jumps allowed in AJD term structure models.

[11]The equivalence between the infinitesimal operator and transition density can be proved by the Hill-Yoshida theorem in Dynkin (1965).

$$M_t^f = f(X_t) - f(X_0) - \int_0^t (\mathcal{A}f)(X_s)ds \qquad (3.31)$$

is a martingale for every $f \in D(\mathcal{A})$, is called a solution to the martingale problem associated with the operator $\mathcal{A}$. This "martingale problem" is a variation of the weak solution to the stochastic differential eqnarray in (3.28) and hence can be employed as the identification condition (Revuz and Yor, 2005).[12]

The martingale property of the transformed processes in (3.31) can be written as a conditional moment restriction:

$$E\left[M_t^f(\theta)|\mathcal{I}_{t'}\right] = M_{t'}^f(\theta)$$

for any $f \in D(\mathcal{A})$ and $t' < t$, where callI$_{t'} = \sigma\{X_{t''}\}_{t''<t'}$ is the information set generated by the history of $\{X_t\}$ at time t'. For the econometric convenience, the following equivalent conditional moment restrictions can be derived by the martingale difference (*m.d.s.*) property of the first-order difference of the transformed process $M_t^f(\theta)$:

$$E\left[Z_t^f(\theta)|\mathcal{I}_{t'}\right] = 0 \qquad (3.32)$$

for any $t' < t$, where callI$_{t'} = \sigma\{X_{t''}\}_{t''<t'}$, $Z_t^f(\theta) = M_t^f(\theta_0) - M_{t-\Delta}^f(\theta)$, and Δ is the sampling interval. By the Markov property, (3.32) is further equivalent to

$$E\left[Z_t^f(\theta)|X_{t-\Delta}\right] = 0 \qquad (3.33)$$

for any $f \in D(\mathcal{A})$, which is the moment condition we shall utilize for proposing the estimator.

[12]There are two types of solutions to a stochastic differential eqnarray, the strong solution and the weak solution. Loosely speaking, the difference between strong and weak solutions, intuitively, is very similar to that between a random variable and its distribution. Since econometric inferences are concerned only with the dynamic probability laws of the process instead of specific sample paths, it is sufficient to consider a weak solution.

Observe that one potential difficulty with the moment condition in (3.33) is that there are an infinite number of functions $f(\cdot)$ in the domain $D(\mathcal{A})$. This is a general problem which also appears in Hansen and Scheinkman (1995), Conley, Hansen, Luttmer and Scheinkman (1997), Kanaya (2007) and Song (2011). To tackle this difficulty, the domain of the operator $\mathcal{A}$ must be reduced. For diffusion models, a subclass of $D(\mathcal{A})$ can be chosen without losing identification information (Song, 2011; Karatzas and Shreve, 1991, Proposition 4.6): $f(x) = x_i$ and $f(x) = x_i x_j$ with $1 \leq i, j \leq d$. This choice yields the moment condition (3.33) with $Z_t^f(\theta)$ as a vector of components for $i, j = 1, \cdots, d$

$$\begin{aligned}
Z_t^i(\theta_0) &= M_t^{x_i}(\theta_0) - M_{t-\Delta}^{x_i}(\theta_0) \\
&= X_t^i - X_{t-\Delta}^i - \int_{t-\Delta}^t b_i(X_s;\theta_0)ds \\
Z_t^{i,i}(\theta_0) &= M_t^{x_i x_i}(\theta_0) - M_{t-\Delta}^{x_i x_i}(\theta_0) \\
&= \left(X_t^i\right)^2 - \left(X_{t-\Delta}^i\right)^2 - \int_{t-\Delta}^t \left[2b_i(X_s;\theta_0)X_s^i + \sum_{k=1}^d \sigma_{i,k}(X_s;\theta_0)^2\right] ds \\
Z_t^{i,j}(\theta_0) &= M_t^{x_i x_j}(\theta_0) - M_{t-\Delta}^{x_i x_j}(\theta_0) \\
&= X_t^i X_t^j - X_{t-\Delta}^i X_{t-\Delta}^j - \int_{t-\Delta}^t \Big[b_i(X_s;\theta_0)X_s^j + b_j(X_s;\theta_0)X_s^i \\
&\quad + \frac{1}{2}\sum_{k=1}^d \sigma_{i,k}(X_s;\theta_0)\sigma_{j,k}(X_s;\theta_0)\Big] ds \qquad (3.34)
\end{aligned}$$

for $i \neq j$.

For the general jump-diffusion model in (3.28), the exponential functions are chosen based on the concept of a core and "approximation" theory (see Chapter 2): $f(x) = \exp\left[-\left(x_1^2 + \cdots + x_d^2\right)/2\right]$. This delivers the conditional moment restriction (3.33) with

$$\begin{aligned}
Z_t^f(\theta) &= e^{-\left(X_{1,t}^2 + \cdots + X_{d,t}^2\right)/2} - e^{-\left(X_{1,t-\Delta}^2 + \cdots + X_{d,t-\Delta}^2\right)/2} \\
&\quad - \int_{t-\Delta}^t \mathcal{A}_\theta e^{-\left(X_{1,s}^2 + \cdots + X_{d,s}^2\right)/2} ds, \qquad (3.35)
\end{aligned}$$

where

$$
\begin{aligned}
& \mathcal{A}_\theta e^{-\left(X_{1,s}^2+\cdots+X_{d,s}^2\right)/2} \\
= \;& e^{-\left(X_{1,s}^2+\cdots+X_{d,s}^2\right)/2}\left\{-\sum_{i=1}^{d} b_i(X_s;\theta)X_{i,s}+\frac{1}{2}\sum_{i,j=1}^{d} a_{i,j}(X_s;\theta)X_{i,s}X_{j,s}\right. \\
& \left.-\frac{1}{2}\sum_{i,j=1}^{d} a_{i,j}(X_s;\theta)+\lambda(X_s,\theta)\int\left[e^{-c\cdot X_s-|c|^2/2}-1\right]d\nu\left(c,\theta\right)\right\}, \qquad (3.36)
\end{aligned}
$$

It can be observed that the conditional moment restrictions are expressed explicitly by the drift, volatility, and jump terms and can be used directly.[13] In contrast, the transition density-based methods as in Lo (1988) and Ait-Sahalia (2002, 2008) have to approximate the transition density or numerically solve it because the transition density rarely has a closed form. The infinitesimal operator methods are particularly convenient for multivariate models, for which the transition density methods are extremely complicated and computationally inconvenient.

Next, an asymptotically efficient estimator will be proposed for the model defined by the following conditional moment condition:

$$E\left[u(X_t, X_{t-\Delta};\theta)|X_{t-\Delta},\cdots,X_{t-m\Delta}\right]=0, \qquad (3.37)$$

where $\{X_t\}_{t=\Delta,2\Delta,\cdots}$ is a strictly stationary and m-th order Markov process. It can be observed that the moment condition (3.33) is a special case with m=1.

Suppose $x_t = (X_{t-\Delta},\cdots,X_{t-m\Delta})'$ and $y_t = (X_t, x_t')'$; then (3.37) can be restated as $E\left[u\left(y_t;\theta\right)|x_t\right] = 0$. The dimensions of the vectors are $dm\times 1$ for x_t, $(d+1)\,m\times 1$ for y_t, and $q\times 1$ for the function $u\left(\cdot;\theta\right)$. To construct the estimator, first define the

[13]Jumps and volatilities can be identified because they are specified separately in the parametric infinitesimal operator, as can be seen from (3.35)-(3.36).

positive weights

$$w_{ij} = \frac{\mathcal{K}\left[\left(x_i - x_j\right)/b_n\right]}{\sum_{j=1}^{n} \mathcal{K}\left[\left(x_i - x_j\right)/b_n\right]} \triangleq \frac{\mathcal{K}_{ij}}{\sum_{j=1}^{n} \mathcal{K}_{ij}}, \tag{3.38}$$

where $\mathcal{K}(\cdot)$ is a kernel function and b_n is a sequence of positive bandwidth numbers. Let p_{ij} be the probability mass placed at $\left(x_i, y_j\right)$ by a discrete distribution supported on $\{x_1, \cdots, x_n\} \times \{y_1, \cdots, y_n\}$, which can be regraded as an estimate of the conditional probability $P\left\{y = y_j | x = x_i\right\}$. We can then form the following maximization problem, which is essentially a "nonparametric maximum likelihood" approach:

$$\begin{aligned} &\max_{p_{ij}} \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} \log p_{ij} \\ s.t. p_{ij} \quad \geq \quad &0, \sum_{j=1}^{n} p_{ij} = 1, and \sum_{j=1}^{n} u\left(y_j; \theta\right) p_{ij} = 0 \end{aligned} \tag{3.39}$$

for $i, j = 1, \cdots, n$.

Problem (3.39) can be conveniently solved by a Lagrangian multiplier method. Let

$$\begin{aligned} \mathcal{L}(\theta) \quad = \quad & \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} \log p_{ij} - \sum_{i=1}^{n} \mu_i \left(\sum_{j=1}^{n} p_{ij} - 1 \right) \\ & - \sum_{i=1}^{n} \lambda_i' \left(\sum_{j=1}^{n} u\left(y_j; \theta\right) p_{ij} \right), \end{aligned} \tag{3.40}$$

where $\mu_1, \cdots, \mu_n$ and $\lambda_1, \cdots, \lambda_n$ are the Lagrangian multipliers for the second and third sets of constraints respectively. By Kitamura, Tripathi and Ahn (2004), the solution is

$$\widehat{p}_{ij} = \frac{w_{ij}}{1 + \lambda_i' u\left(y_j; \theta\right)}, \tag{3.41}$$

where

$$\sum_{j=1}^{n} \frac{w_{ij} u\left(y_j; \theta\right)}{1 + \lambda_i' u\left(y_j; \theta\right)} = 0 \tag{3.42}$$

for each θ and $i = 1, \cdots, n$. Now we can :form the local empirical log-likelihood (LELL) function at θ and then define the local empirical likelihood estimator for the model (3.37) as

$$\begin{aligned} \widehat{\theta}_{LEL} &= \arg\max_{\theta\in\Theta} \left\{ \mathbf{LELL}(\theta) = \sum_{i=1}^{n}\sum_{j=1}^{n} T_{i,n} w_{ij} \log \widehat{p}_{ij} \right. \\ &= \left. \sum_{i=1}^{n}\sum_{j=1}^{n} T_{i,n} w_{ij} \log \left[\frac{w_{ij}}{1 + \lambda_i' u\left(y_j; \theta\right)} \right] \right\}, \end{aligned} \tag{3.43}$$

where λ_i solves (3.42) and $T_{i,n} \equiv 1\left\{\widehat{h}(x_i) > b_n^{\varsigma}\right\}$ with $\varsigma \in (0,1)$ is a sequence of trimming functions (see Chapter 2 for details about $T_{i,n}$). Chapter 2 shows that the estimator $\widehat{\theta}_{LEL}$ is consistent and asymptotically normal and provides a consistent covariance matrix estimator. Moreover, it achieves the semi-parametric efficiency bound in Carrasco and Florens (2008) and hence is asymptotically efficient. See Chapter 2 for details about the asymptotic properties of $\widehat{\theta}_{LEL}$.

3.3.2 AJD Term Structure Models with Latent Factors

Observe that, in the AJD term structure model introduced in Section 3, the state variable X_t is not observable. Following the literature (Duffee, 2002; Ait-Sahalia and Kimmel, 2010), the values of the state vector X_t can be extracted from a set of yields due to the affine structure of the model. Since the number of observed yields is usually larger than the number of state variables and some of the yields can be written as deterministic functions of other observed yields without error, choices need be made as to which yields to use in extracting the state vector X_t. By the standard treatment in the literature (Dai and Singleton, 2000, 2002, 2003; Duffee, 2002; Ait-Sahalia and Kimmel, 2010), certain benchmark yields are assumed to be observed precisely while other yields are observed with measurement errors. By further assuming these measurement errors to be *i.i.d*. and independent of the state variables (Duffee, 2002; Ait-Sahalia and Kimmel, 2010), an additional moment condition is obtained and combined together with the moment condition from the state variable dynamics to form the basis of the estimation. In the following, I elaborate on the details of the estimation procedures.

The first task for estimating AJD term structure models is to infer the state variable X_t, which is not directly observable, from the cross-section of bond yields at date t with various maturities. This inversion is implementable since yields of zero coupon bonds are affine functions of the state variables, as can be seen from (3.12). To ensure the identification of parameters in the market prices of risks, the employed number of observed yields needs to be larger than the number of state variables since some of the yields can be written as deterministic functions of other observed yields without error. Following the standard treatment in the literature (Duffee, 2002), we we use $d + H$ observed yields ($H > 0$) for the $AJD_M(d)$ model, which include d yields observed precisely with periods to maturity of $\tau_1, \cdots, \tau_d$ and the other H observed with errors with periods to maturity of $\tau_{d+1}, \cdots, \tau_{d+H}$. At each date t, we can obtain X_t exactly by the yields observed without errors based on the following eqnarray:

$$\begin{pmatrix} y(X_t, \tau_1) \\ \vdots \\ y(X_t, \tau_d) \end{pmatrix} = -\begin{pmatrix} A(\tau_1)/\tau_1 \\ \vdots \\ A(\tau_d)/\tau_d \end{pmatrix} + \begin{pmatrix} B(\tau_1)'/\tau_1 \\ \vdots \\ B(\tau_d)'/\tau_d \end{pmatrix}\begin{pmatrix} X_{1t} \\ \vdots \\ X_{dt} \end{pmatrix}, \tag{3.44}$$

which follows from (3.36) with $y(X_t, \tau_1), \cdots, y(X_t, \tau_d)$ comprising the yields having periods to maturity of $\tau_1, \cdots, \tau_d$. Therefore, current values of the state vector

X_t can be obtained by inverting this affine equation:

$$\begin{pmatrix} X_{1t} \\ \vdots \\ X_{dt} \end{pmatrix} = \begin{pmatrix} B(\tau_1)'/\tau_1 \\ \vdots \\ B(\tau_d)'/\tau_d \end{pmatrix}^{-1} \times \left[\begin{pmatrix} y(X_t,\tau_1) \\ \vdots \\ y(X_t,\tau_d) \end{pmatrix} + \begin{pmatrix} A(\tau_1)/\tau_1 \\ \vdots \\ A(\tau_d)/\tau_d \end{pmatrix} \right]. \tag{3.45}$$

Observed that these extracted state variables X_t are expressed in terms of benchmark yields $y(X_t,\tau_1)$, $\cdots$, $y(X_t,\tau_d)$ and coefficients $A(\tau_1),\cdots,A(\tau_d)$ and $B(\tau_1),\cdots,B(\tau_d)$, which in fact are functions of the parameter vector θ. By the infinitesimal operator-based procedures specified in Section 3.3.1, we obtain the moment condition characterizing the dynamics of the model:

$$E\left[Z_t^f(\theta)|X_{t-\Delta}\right] = 0 \tag{3.46}$$

for some $\theta_0 \in \Theta$, where $Z_t^f(\theta)$ is defined as in (3.33)-(3.36). (3.46) is the identification condition for the AJD term structure model we consider when only the benchmark yields are employed.

Besides the benchmark yields, we have yields with periods to maturity of $\tau_{d+1},\cdots,\tau_{d+H}$, which are assumed to be observed with errors. By the extracted state variables X_t and the coefficients $A(\tau_1),\cdots,A(\tau_d)$ and $B(\tau_1),\cdots,B(\tau_d)$ in (3.44), we can calculate the implied values of these H yields as follows:

$$\begin{pmatrix} y(X_t,\tau_{d+1}) \\ \vdots \\ y(X_t,\tau_{d+H}) \end{pmatrix} = -\begin{pmatrix} A(\tau_{d+1})/\tau_{d+1} \\ \vdots \\ A(\tau_{d+H})/\tau_{d+H} \end{pmatrix}$$

$$+\begin{pmatrix} B(\tau_{d+1})'/\tau_{d+1} \\ \vdots \\ B(\tau_{d+H})'/\tau_{d+H} \end{pmatrix}\begin{pmatrix} X_{1t} \\ \vdots \\ X_{dt} \end{pmatrix}, \tag{3.47}$$

The observation errors are equal to the differences between these implied yields and the observed yields from the data, denoted as $\widehat{y}(X_t, \tau_{d+1}), \cdots, \widehat{y}(X_t, \tau_{d+H})$:

$$\begin{pmatrix} \epsilon_t(\tau_{d+1}) \\ \vdots \\ \epsilon_t(\tau_{d+H}) \end{pmatrix} = \begin{pmatrix} \widehat{y}(X_t, \tau_{d+1}) \\ \vdots \\ \widehat{y}(X_t, \tau_{d+H}) \end{pmatrix} - \begin{pmatrix} Y(X_t, \tau_{d+1}) \\ \vdots \\ Y(X_t, \tau_{d+H}) \end{pmatrix}, \tag{3.48}$$

By standard assumptions in the literature (Duffee, 2002; Ait-Sahalia and Kimmel, 2010), these errors have zero conditional mean and constant conditional variances, are independent across time and maturity, and are also independent of the state variables. Consequently we have

$$E\left[\left.\begin{pmatrix} \epsilon_t(\tau_{d+1};\theta) \\ \cdots \\ \epsilon_t(\tau_{d+H};\theta) \end{pmatrix}\right| X_{t-\Delta}\right] = 0, \tag{3.49}$$

where $\epsilon_t(\tau_{d+1}), \cdots, \epsilon_t(\tau_{d+H})$ depend on the parameter vector θ by (3.47) and (3.48). Denote $\epsilon_t(\theta)$ the H-dimensional vector with components $\epsilon_t(\tau_{d+1};\theta), \cdots, \epsilon_t(\tau_{d+H};\theta)$. Then (3.49) delivers

$$E[\epsilon_t(\theta)|X_{t-\Delta}] = 0 \tag{3.50}$$

for the θ_0 in (3.46), which we can take as the additional identification condition related to the yields observed with errors. Compared with Ait-Sahalia and Kimmel (2010), who assume that the errors have a Gaussian distribution to obtain the likelihood functions, the identification condition (3.50) employed here is robust and free of mis-specifications of distributions. Finally, conditions (3.46) and

(3.50) are combined to form an augmented identification condition for the AJD term structure model:

$$E\left[\left.\begin{pmatrix} Z_t^f(\theta) \\ \epsilon_t(\theta) \end{pmatrix}\right| X_{t-\Delta}\right] = 0, \theta_0 \in \Theta. \tag{3.51}$$

The condition (3.51) is our essential basis in the empirical studies of a specific AJD term structure model. The proposed **LEL** approach in Section 3.3.1 is then applied to (3.51) for estimating the model.

3.4 Empirical Performances of AJD Models

In this section, I evaluate the empirical performance of AJD term structure models in capturing time variations in risk premiums and conditional volatilities.[14] Following the literature (Dai and Singleton, 2000, 2002, 2003; Duffee, 2002; Litterman and Scheinkman, 1991), three-factor essentially AD and AJD models are considered. In fact, the principal component analysis in Panel C of Table 3.1 shows that the first three principal components can explain more than 99.9% of yields variations, which provides a justification for using three-factor models. I shall first discuss an additional criterion, the risk-premium adjusted projections advocated by Dai and Singleton (2002), that can be used to match the time-varying risk premium as a complement to the "yield regression" studied in Section 2. Then the evaluation procedure is presented for investigating the empirical performance of the models.

[14] I emphasize that this is an effort undertaken to check empirically whether a dynamic term structure model has quantitative implications that are consistent with the data. In contrast, many studies have examined the relationship between theoretical term structure models and the "expectation hypothesis" (Cox, Ingersoll and Ross, 1981; Jarrow, 1981; Campbell, 1986; Longstaff, 2000).

3.4.1 Risk-Premium Adjusted Projections

As Dai and Singleton (2002) argue, if the empirical failure of the ″expectation hypothesis″ documented in Section 3.1 is due to time-varying risk premiums, accommodating risk premiums in the ″yield regression″ should make it possible to restore a slope coefficient of one. In particular, (3.2) implies that

$$E_t\left[y_{t+1}^{\tau-1} - y_t^{\tau}\right] + e_t^{\tau} / (\tau - 1) = (y_t^{\tau} - r_t) / (\tau - 1)\,.$$

It follows that, once we adjust the yield changes $y_{t+1}^{\tau-1} - y_t^{\tau}$ by $e_t^{\tau} / (\tau - 1)$, we should recover the coefficient of unity desired by the advocates of the ″expectation hypothesis″ in the following ″risk-adjusted yield regression″ (Dai and Singleton, 2002):

$$y_{t+1}^{(\tau-1)} - y_t^{\tau} + \frac{e_t^{\tau}}{\tau - 1} = constant + \phi_{\tau T}^{R}\left(\frac{y_t^{\tau} - r_t}{\tau - 1}\right) + residual. \tag{3.52}$$

The sample coefficients $\phi_{\tau T}^{R}$, obtained using historical yields for y_t^{τ} and model-fitted risk premiums in constructing e_t^{τ}, will be closed to unity only when the model accurately captures the dynamics of risk premiums–in other words, only when the model matches the term structure dynamics under the risk-neutral measure Q (Dai and Singleton, 2002). This is termed **LPY (ii)** in Dai and Singleton (2002). In contrast, matching the empirical failure pattern of the ″yield regression″ points to the historical behavior of the term structure dynamics under the physical measure $\mathcal{P}$. These two properties are not equivalent and matching both should be the criterion for determining whether dynamic term structure models successfully capture the time variation in risk premiums (Dai and Singleton, 2002).

3.4.2 Evaluation Procedures of Model Performances

To evaluate the empirical performances of affine term structure models, I first estimate all four subclasses of essentially AD models ($EA_0(3)$, $EA_1(3)$, $EA_2(3)$, and $EA_3(3)$) and AJD models ($EAJD_0(3)$, $EAJD_1(3)$, $EAJD_2(3)$, and $EAJD_3(3)$), respectively, using the infinitesimal operator methods proposed in Section 3.3. Explanation of the details of the estimation results, reported in Tables 3.4-3.5, are deferred to Section 3.5 with the analysis of jump risk premiums.

From all the discussions in Sections 3.1 and 3.4.1, there are two components needed for evaluating the empirical performance of the models in matching time-varying risk premiums: the empirical pattern of the "yield regression" (3.3) documented in Figure 3.2 and the unity line of the "risk-adjusted yield regression" (3.52). For the former, I follow Dai and Singleton (2002) to compare the model-implied population coefficients

$$\phi_\tau = \frac{cov\left(y_{t+1}^{(\tau-1)} - y_t^\tau,\ (y_t^\tau - r_t) / (\tau - 1)\right)}{var\left((y_t^\tau - r_t) / (\tau - 1)\right)} \tag{3.53}$$

with their sample counterparts in Figure 3.2[15]. These population coefficients are calculated by treating the model parameter estimates as the true values and then applying the analytic formulas to compute the moments in (3.53). The yields data do not enter (3.53) directly; they show up only through the parameter estimates. For the latter, the sample counterparts $\phi_{\tau T}^R$ of the population coefficients

$$\phi_\tau^R = \frac{cov\left(y_{t+1}^{(\tau-1)} - y_t^\tau + e_t^\tau / (\tau - 1)\,,\ (y_t^\tau - r_t) / (\tau - 1)\right)}{var\left((y_t^\tau - r_t) / (\tau - 1)\right)} \tag{3.54}$$

[15]Dai and Singleton (2002) find that the model-implied ϕ_τ, computed from the sample of model-implied fitted yields (obtained by inverting the model for the fitted-state variables, computing model-implied fitted zero-coupon bond yields, and then estimating the "yield regression" in (3.3)), can give very misleading conclusions pertaining to the actual popluation distributions implied by the models. The reason for this is that it mixes the properties of a dynamic term structure model with those of the historical data.

will be examined to see if they are statistically different from a horizontal line at one which is the model-implied value of ϕ_τ^R. Here the historical yields y_t^τ will be used while e_t^τ are calculated using the fitted state variables.

To evaluate the empirical performance of the models in matching the time-varying conditional volatility, the sample variances, computed using the simulated time series of LIBOR-Swap rates from the models evaluated at their estimated parameter values, are compared with the empirical pattern in Figure 3.3. Two aspects in the comparisons are considered: the hump shape and the levels of volatilities (Dai and Singleton, 2000; Piazzesi, 2005). Furthermore, I follow Dai and Singleton (2003) and Buraschi, Cieslak and Trojani (2008) to estimate a GARCH(1,1) model for the simulated yields from different term structure models and check the degree of model-implied time-varying volatility relative to what is found in Table 3.3 using historical data. In the simulations, 18 years of daily data are generated to be comparable to the historical data sample. The parameter estimates for AD and AJD models are computed as follows: First, 1000 sample paths of the yields of a specific maturity are simulated from the models; second, the GARCH(1,1) model is estimated for each of the sample path; finally, the median of these 1000 parameter estimates are taken as the model-implied coefficients.

3.4.3 Performance of Three-Factor AD Models

Similar to the results found by Dai and Singleton (2002) for U.S. Treasury yields, I shall investigate whether the three-factor essentially AD term structure models can capture time variations in both the risk premium and conditional volatility.

Then the model structures will be analyzed to provide intuitive theoretical support for the empirical successes and failures.

We discuss the empirical results first on time-varying risk premium and then on time-varying conditional volatility.

Time-Varying Risk Premiums. Based on the parameter estimates in Table 3.4 for three-factor essentially AD term structure models, the population coeffici-

Table 3.4: **Parameter Estimates for Three-Factor Essentially AD Term Structure Models**

	$EA_0(3)$		$EA_1(3)$		$EA_2(3)$		$EA_3(3)$	
	Est.	SE	Est.	SE	Est.	SE	Est.	SE
κ_{11}	0.0420	(0.0281)	0.0967	(0.0052)	0.1154	(0.1332)	0.3166	(0.0667)
κ_{12}	0.0	-	0.0	-	1.0458	(0.4007)	0.0023	(0.3009)
κ_{13}	0.0	-	0.0	-	0.0	-	-0.0222	(0.0145)
κ_{21}	-0.1513	(0.2444)	-0.0684	(0.0152)	-0.1150	(0.0726)	0.0440	(0.0213)
κ_{22}	0.0371	(0.0026)	0.1273	(0.0904)	0.7451	(0.1111)	0.0775	(0.0074)
κ_{23}	0.0	-	-3.1449	(1.4334)	0.0	-	0.0702	(0.0667)
κ_{31}	3.3572	(2.4403)	0.0128	(0.0206)	0.1485	(0.2234)	-0.0463	(0.0061)
κ_{32}	-0.0193	(0.0804)	0.3532	(0.0925)	0.0933	(0.0458)	0.0208	(0.0092)
κ_{33}	0.0415	(0.0140)	0.1014	(0.1307)	0.1652	(0.0149)	0.2771	(0.1709)
θ_1	0.0	-	0.1067	(0.0885)	0.0955	(0.0210)	0.0709	(0.0167)
θ_2	0.0	-	0.0	-	0.2202	(0.2911)	0.0566	(0.0320)
θ_3	0.0	-	0.0	-	0.0	-	0.0622	(0.0384)
α_1	1.0	-	0.0	-	0.0	-	0.0	-
α_2	1.0	-	1.0	-	0.0	-	0.0	-

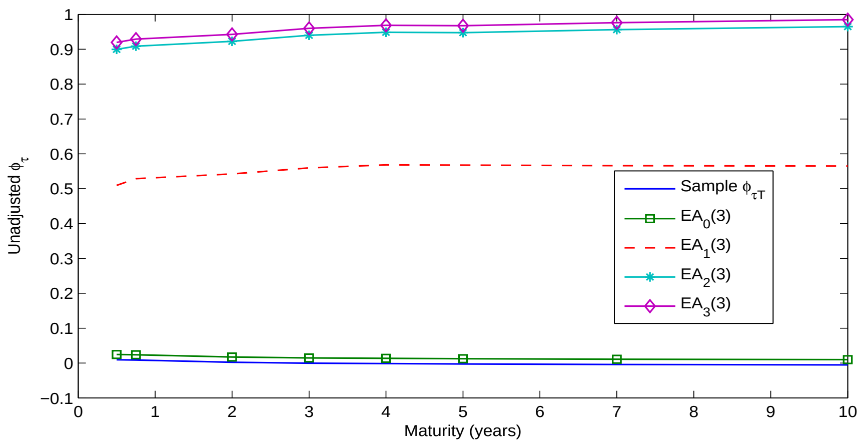
α_3	1.0	-	1.0	-	1.0	-	0.0	-
β_{11}	0.0	-	1.0	-	1.0	-	1.0	-
β_{12}	0.0	-	0.0	-	0.0	-	0.0	-
β_{13}	0.0	-	0.0	-	0.0	-	0.0	-
β_{21}	0.0	-	0.1635	(0.0291)	0.0	-	0.0	-
β_{22}	0.0	-	0.0	-	1.0	-	1.0	-
β_{23}	0.0	-	0.0	-	0.0	-	0.0	-
β_{31}	0.0	-	0.2991	(0.2507)	0.1212	(0.1404)	0.0	-
β_{32}	0.0	-	0.0	-	0.0367	(0.0403)	0.0	-
β_{33}	0.0	-	0.0	-	0.0	-	1.0	-
η_{11}	-0.3008	(0.0042)	-0.1544	(0.2033)	-0.2247	(0.2481)	-0.0190	(0.0655)
η_{12}	-0.0213	(0.0114)	-0.2007	(0.2011)	-0.0913	(0.0449)	-0.0418	(0.0248)
η_{13}	-0.2822	(0.0211)	-0.0688	(0.0443)	-0.0701	(0.0545)	-0.1406	(0.0221)
δ_0	-0.0226	(0.0040)	0.2410	(0.0748)	-0.0416	(0.0903)	0.6004	(0.0811)
δ_{11}	0.0690	(0.0433)	0.1152	(0.1067)	0.0707	(0.0242)	0.7107	(2.1633)
δ_{12}	0.0172	(0.0089)	0.0934	(0.0805)	0.8805	(0.1924)	0.2022	(0.2139)
δ_{13}	0.1006	(0.0088)	0.3444	(0.0455)	-0.1430	(0.0782)	0.0467	(0.0141)
$\eta_{2,11}$	-0.0122	(0.0227)	0.0	-	0.0	-	0.0	-
$\eta_{2,12}$	-0.1709	(0.1066)	0.0	-	0.0	-	0.0	-
$\eta_{2,13}$	-0.5282	(0.1544)	0.0	-	0.0	-	0.0	-
$\eta_{2,21}$	-0.4111	(0.2554)	-0.0157	(0.0266)	0.0	-	0.0	-
$\eta_{2,22}$	-0.2333	(0.2607)	-0.0333	(0.0182)	0.0	-	0.0	-
$\eta_{2,23}$	-0.2900	(0.0177)	-0.0910	(0.0187)	0.0	-	0.0	-
$\eta_{2,31}$	-0.3171	(0.2403)	-0.0229	(0.0316)	-0.0114	(0.2548)	0.0	-
$\eta_{2,32}$	-0.1155	(0.0778)	-0.0142	(0.0040)	-0.0502	(0.0133)	0.0	-
$\eta_{2,33}$	-0.1582	(0.1267)	-0.0143	(0.0122)	-0.0704	(0.0166)	0.0	-

This table reports parameter estimates for three-factor essentially AD models using LIBOR-Swap rates with maturities of 6-month, 2-year, 10-year, 3-year, 5-year, and 7-year and sampled daily from August 13, 1990 to December 31, 2008. The rates with the first three maturities are assumed to be observed without error and the remaining three observed with error. Estimation follows the infinitesimal operator based procedure described in Section 4.2. Specifications of models, $EA_0(3)$, $EA_1(3)$, $EA_2(3)$ and $EA_3(3)$, have the sport rate as $r_t = \delta_0 + \delta_{11} X_{1t} + \delta_{12} X_{2t} + \delta_{13} X_{3t}$, risk-neutral dynamics for X_t in (3.18), (3.20), (3.22) and (3.24) and physical dynamics for X_t in (3.17), (3.19), (3.21) and (3.23) with the jump terms eliminated. Reported in the parentheses of the "SE" column are the standard errors of the coefficients estimates. The blanks in the "SE" columns refer to those parameters pre-specified by the model structures.

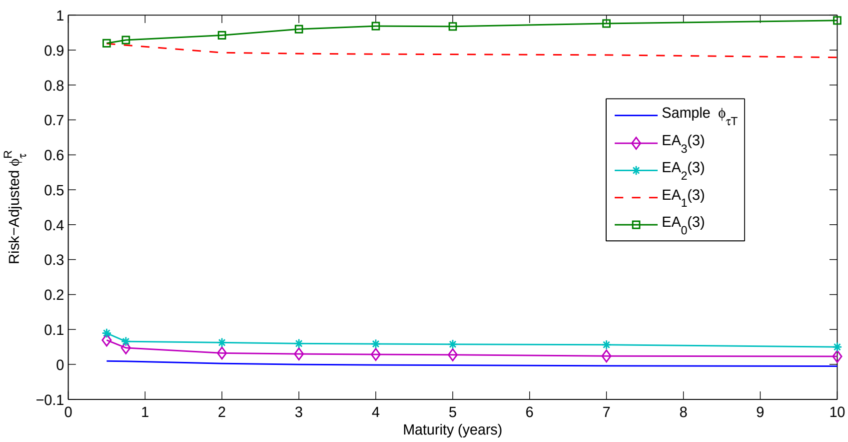
-ents ϕ_τ in (3.53) for all four subclasses ($EA_0(3)$, $EA_1(3)$, $EA_2(3)$, and $EA_3(3)$) are displayed in Figure 3.4.A, together with the historical $\phi_{\tau T}$ from Table 3.2. We see that only the coefficients from the $EA_0(3)$ model closely resemble the historical pattern. None of the other three ($EA_1(3)$, $EA_2(3)$, and $EA_3(3)$) performs nearly as well at replicating the empirical failure pattern of the "yield regression" in the population; note in particular that the $EA_2(3)$ and $EA_3(3)$ models, with relatively flexible time-varying conditional volatility specifications, are approximately horizontal lines at unity. Moreover, these two models have in fact an upward sloping pattern for the population coefficients, which is in sharp contrast to the downward sloping pattern of the historical projection coefficients. From these comparisons, I conclude that, similarly to Dai and Singleton (2002) for U.S. Treasury yields, only the model $EA_0(3)$ is successful at matching the time variation in risk premiums represented by the empirical failure pattern of

Figure 3.4: **AD Model-Implied Patterns for Time-Varying Risk Premiums**

This figure reports values of slope coefficients ϕ_τ and ϕ^R_τ from the "yield regression" of $y^{(\tau-1)}_{t+1} - y^\tau_t$ onto $(y^\tau_t - r_t)\ /\ (\tau - 1)$, and "risk-adjusted yield regression" of $y^{(\tau-1)}_{t+1} - y^\tau_t + e^\tau_t /\ (\tau - 1)$ onto $(y^\tau_t - r_t)\ /\ (\tau - 1)$ respectively, for the $EA_0(3)$, $EA_1(3)$, $EA_2(3)$, and $EA_3(3)$ models. The population coefficients are obtained treating the parameter estimates as the true values and then applying analytic formulas to compute the involved moments in (3.53). "Sample $\phi_{\tau T}$" displays the estimated coefficients reported in Table 3.2.



(a) Model-Implied Patterns for the "Yield Regression"



(b) Model Implied Patterns for the "Risk-Adjusted Yield Regression" the "yield regression."

We turn now to the "risk-adjusted yield regression," which characterizes whether or not a dynamic term structure model can capture term structure dynamics under the risk-neutral measure. It is expected that the term structure of risk-adjusted ϕ_τ^R should be a horizontal line at unity if the risk premiums are well-specified. Figure 3.4.B presents the model-implied ϕ_τ^R along with the historical projection coefficients from Table 3.2. It is observed that the $EA_2(3)$ and $EA_3(3)$ models fail completely to match the unity line. In fact, adjusting for risk premiums in these models produces almost the same results as in the unadjusted regressions results in Table 3.2; that is, both $EA_2(3)$ and $EA_3(3)$ models perform as if they have constant risk premiums. In contrast, the risk-adjusted coefficients ϕ_τ^R for both the $EA_0(3)$ and $EA_1(3)$ models lie close to unity, with the former performing better. However, the $EA_1(3)$ model has a tendency to deviate from the unity line as maturity τ increases. For example, when $\tau = 10$ (years), the risk-adjusted coefficient ϕ_τ^R is smaller than 0.9. Hence, differing from Dai and Singleton's (2002) finding using U.S. Treasury yields that both the $EA_0(3)$ and $EA_1(3)$ models meet the challenge of matching term structure dynamics under the risk-neutral measure, I find here that only the $EA_0(3)$ model can capture the risk-neutral dynamics of the LIBOR-Swap yields.

Summarizing the empirical performance above of three-factor affine models, I find that only the $EA_0(3)$ model in the three-factor essentially AD class is able to capture time variations in the risk premiums for the LIBOR-Swap rates. Another observation is that even for the $EA_0(3)$ model, Figure 3.4.B shows a gap between the risk-adjusted coefficient ϕ_τ^R and unity for the LIBOR-Swap rates with maturities under two years. I shall investigate this issue on the short end of the term structure movements in Section 5.5 in the context of the AJD framework.

Time-Varying Conditional Volatilities. For the volatility persistence, Table 3.5 reports parameter estimates of the GARCH(1,1) model for both the historical and simulated 5-year yield. Similar to Dai and Singleton (2003) and Buraschi, Cieslak and Trojani (2008), the choice of the 5-year yield is motivated by the fact that it is not involved in the estimation of the models[16]. Expectely, we can see that the model $EA_0(3)$, which assums constant volatility functions for the state variables, exhibits little volatility persistence. The model $EA_1(3)$, while understating the degree of volatility persistence a bit, performs much better than $EA_0(3)$ due to the one factor entering the volatility function. The implied GARCH(1,1) estimates form models $EA_2(3)$ and $EA_3(3)$ both match those in the sample quite closely.

For the humped-shape of the volatility term structure, Figure 3.5 plots the sample variances computed using simulated time series of LIBOR-Swap rates from the models evaluated at their estimated parameter values. Observe that, except for the $EA_3(3)$ model, the other three subclasses exhibit a hump. The humps of both $EA_0(3)$ and $EA_1(3)$ coincide with the historical pattern while that of $EA_2(3)$ occurs at around the 2-year maturity. The magnitude of sample variances for the $EA_0(3)$ model is too small to match that of the historical variances. Summarizing the results about both the humped volatility term structure and the high degree of volatility persistence, the conclusion is that $EA_1(3)$ is the best model to capture time variations in the conditional volatility of LIBOR-Swap yields, similar to Dai and Singleton (2000).

[16]I actually made comparisons of models using yields of all available maturities. Results do not change by different choices of yields and hence are omitted here to save space.

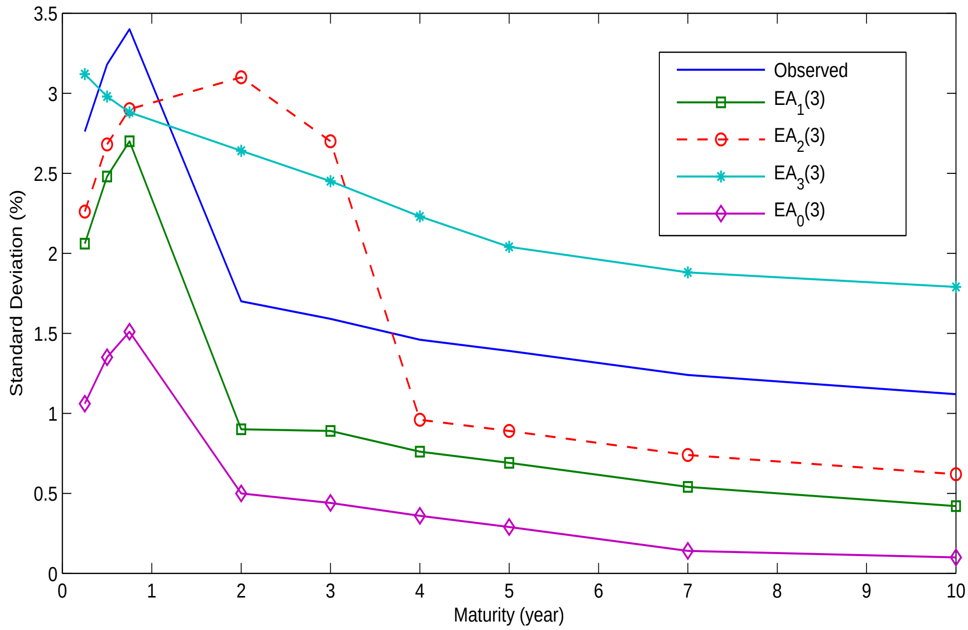
Table 3.5: **GARCH(1,1) Parameters for the Model-Implied LIBOR-Swap Yields**

This table reports the maximum likelihood estimates of a GARCH(1,1) model ($\sigma_t^2 = \overline{\sigma} + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2$, where ε_t is the innovation from the AR(1) representation of the level of the yield) using 5-year yields simulated from eight AD and AJD term structure models. The "data" row presents the parameter estimates in Table 3.3 using sample data of the 5-year swap yield. The parameter estimates for AD and AJD models are computed as follows: First, 1000 sample paths of the 5-year yields are simulated from the models; second, the GARCH(1,1) model is estimated for each of the sample path; finally, the median of these 1000 parameter estimates are taken as the estimates of the model.

	$\overline{\sigma}$	α	β
Data	0.0001	0.1679	0.7780
AD Models			
$EA_0(3)$	0.0008	0.4223	0.0382
$EA_1(3)$	0.0003	0.1755	0.6906
$EA_2(3)$	0.0002	0.1702	0.7311
$EA_3(3)$	0.0001	0.1660	0.8033
AJD Models			
$EAJD_0(3)$	0.0007	0.4305	0.0349
$EAJD_1(3)$	0.0002	0.1682	0.7475
$EAJD_2(3)$	0.0001	0.1657	0.7815
$EAJD_3(3)$	0.0001	0.1608	0.8118

Figure 3.5: **The AD Model-Implied Term Structure of Volatilities**

This figure plots the term structure of the sample variances, computed using simulated time series of daily changes in the logarithms of LIBOR-Swap rates from the four three-factor essentially AD models ($EA_0(3)$, $EA_1(3)$, $EA_2(3)$, and $EA_3(3)$) evaluated at their estimated parameter values. The "observed" line refers to those computed by the historical rates.



Overall, the empirical finding here is that for the LIBOR-Swap curve, $EA_0(3)$ is the only model closely matching time variations in the risk premium. However, it does not generate enough time variation and persistence in the conditional volatility and does not fit the historical volatility structure of LIBOR-Swap rates very closely. The results for the LIBOR-Swap curve, which are similar to those of Dai and Singleton (2002) for U.S. Treasury yields, show empirically that AD term structure models cannot simultaneously capture time variations in the risk premium and conditional volatility. This can be understood theoretically by analyzing the structures of AD models, specifically the relationship between the market prices of risk and the time-varying conditional volatility of the risk factors, for which some initial clues have been given in Section 3.2.2.

By (3.14), (3.15) and specifications of three-factor essentially affine models in Section 3.2.3, we can see that the market prices of risk in the $EA_m(3)$ model are specified as $\Lambda_t = \sqrt{S_t}\eta_1 + \sqrt{S_t^-}\eta_2 X_t$ where the diagonal matrix S_t^- has zeros in its first m diagonal entries and $\left(\alpha_i + \beta_i' X_t\right)^{-1}$ in entries i=m+1,$\cdots$,3 under the precondition that $\inf\left(\alpha_i + \eta_i' X_t\right) > 0$. In the extreme case with m=3, which delivers the $EA_3(3)$ model, the market prices of risk are $\Lambda_t = \sqrt{S_t}\eta_1$, which coincides with the completely affine specification of Dai and Singleton (2000) and which is very restrictive for the term premiums in that Λ_t is proportional to $\sqrt{S_t}$ and the sign of any element cannot change.

For AD models with m=0, 1, $\cdots$, 2, Λ_t becomes more flexible since the last (3-m) elements can depend on X_t directly and hence are able to switch signs. The smaller m is, the more elements of Λ_t can change sign and the more flexible the market prices of risks are. However, by Dai and Singleton (2000), m is the number of independent Brownian motions driving the time-varying conditional volatility specification of the state-variable dynamics. The analysis in Section 3.2.2 about the affine term structure models tell us that the smaller m is, the more restrictive the time-variation of the conditional volatility is. In the extreme case of m=0, the volatility functions are specified as constants.

Take the $EA_1(3)$ and $EA_2(3)$ models as examples. The former has X_{1t} driving the conditional volatility while the latter has both X_{1t} and X_{2t}. But for the market prices of risks, two elements of the former can depend on X_t directly and only one of the latter can switch signs. Therefore, there is a trade-off between the flexibility of the market prices of risk and the time-varying conditional volatility. The AD term structure models allow for sign switching in the risk premiums only at the expense of limiting time variations in the conditional volatility dynamics. As illustrated in Section 3.2.2 (see also Dai and Singleton (2002), Duffee (2002) and Duarte (2004)), the match of time-varying risk premiums by an affine term structure model depends crucially on the specification of Λ_t. This explains exactly the empirical evidence in Section 3.4.3 that only the model $EA_0(3)$, which does not generate any time-variation in the conditional volatility, can match time-varying risk premiums closely.

Another observation about the affine term structure model which I make here but leave for use in Section 3.4.4.2 is that m also controls the flexibility of the conditional correlation structure. It can be observed that the bigger m is, the more restrictive is the conditional correlation. For example, in the $EA_1(3)$ model, X_{1t} enters the volatility functions of both X_{2t} and X_{3t} while in the $EA_2(3)$ model, the first two risk factors enter the volatility function of the third. Therefore, similar to the market prices of risks, the conditional correlation runs in the opposite direction from the conditional volatility and is important for a flexible time-varying risk premium specification as well.

3.4.4 Performance of Three-Factor AJD Models

In this section, I investigate whether three-factor essentially AJD term structure models can simultaneously capture time variations in the risk premium and conditional volatility accurately for the LIBOR-Swap yields. First, the empirical evidence that two models in the three-factor essentially AJD class can match both closely is presented by repeating the same exercises as those conducted in the last section for AD models. Second, by analyzing the structures of essentially AJD models, I show that the jump risk premium is the key to such empirical successes.

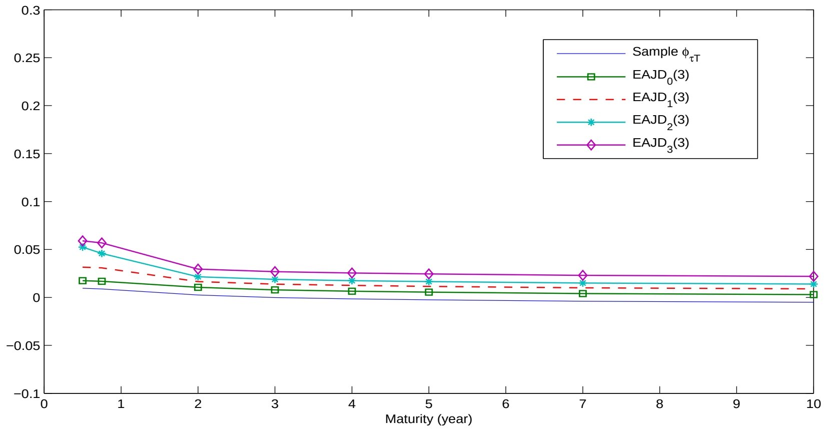
We discuss the empirical results first on time-varying risk premium and then on time-varying conditional volatility.

Time-Varying Risk Premiums. Based on the parameter estimates in Table 3.6 for three-factor essentially AJD term structure models, the population coefficients ϕ_τ in (3.53) for all four models ($EAJD_0(3)$, $EAJD_1(3)$, $EAJD_2(3)$, and $EAJD_3(3)$) are displayed in Figure 3.6.A, together with the historical $\phi_{\tau T}$ from Table 3.2. It is observed that the projection coefficients from all four models closely conform to the historical pattern, with the ranking $EAJD_0(3)$¿$EAJD_1(3)$¿$EAJD_2(3)$¿$EAJD_3(3)$ in terms of their performances. This is very different from the results for the three-factor essentially AD models in Figure 3.4.A using the LIBOR-Swap rates, and obtained in Dai and Singleton (2002) using U.S. Treasury yields, in which only the the $EA_0(3)$ model successfully captures the time variation in risk premiums.

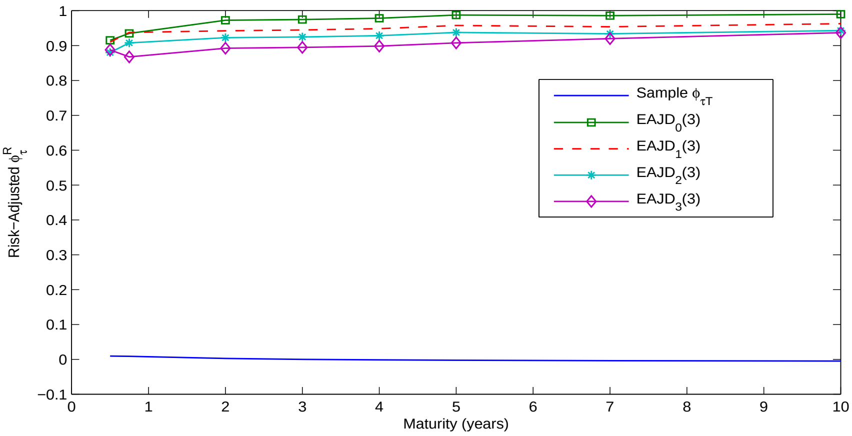
Turning to the ”risk-adjusted yield regression,” Figure 3.6.B presents the model-implied ϕ_τ^R along with the historical projection coefficients from Table 3.2. We can see that, adjusting for risk premiums, the regression coefficients ϕ_τ^R from all four models in the three-factor essentially AJD class are close to unity, with the same rank order as that for the ”yield regression” in terms of the performances, i.e., $EAJD_0(3)$¿$EAJD_1(3)$¿$EAJD_2(3)$¿$EAJD_3(3)$. This also differs from the results obtained for AD models in Figure 3.4.B using the LIBOR-Swap rates (only the $EA_0(3)$ model matches the unity line in the ”risk-adjusted yield reg-

Figure 3.6: **AJD Model-Implied Patterns for Time-Varying Risk Premiums**

This figure reports values of slope coefficients ϕ_τ and ϕ_τ^R from the "yield regression" of $y_{t+1}^{(\tau-1)} - y_t^\tau$ onto $(y_t^\tau - r_t)\ /\ (\tau-1)$, and "risk-adjusted yield regression" of $y_{t+1}^{(\tau-1)} - y_t^\tau + e_t^\tau /\ (\tau-1)$ onto $(y_t^\tau - r_t)\ /\ (\tau-1)$ respectively, for the EAJD$_0$(3), EAJD$_1$(3), EAJD$_2$(3), and EAJD$_3$(3) models. The population coefficients are obtained treating the parameter estimates as the true values and then applying analytic formulas to compute the involved moments in (3.53). "Sample $\phi_{\tau T}$" displays the estimated coefficients reported in Table 3.2.



(a) Model-Implied Patterns for the "Yield Regression"



(b) Model Implied Patterns for the "Risk-Adjusted Yield Regression"

Table 3.6: **Parameter Estimates for Three-Factor Essentially AJD Term Structure Models**

	$EAJD_0(3)$		$EAJD_1(3)$		$EAJD_2(3)$		$EAJD_3(3)$	
	Est.	SE	Est.	SE	Est.	SE		
κ_{11}	0.3130	(0.2228)	0.0130	(0.0012)	0.1880	(0.1044)	0.3007	(0.0894)
κ_{12}	0.0	-	0.0	-	0.1672	(0.0033)	0.0022	(0.0006)
κ_{13}	0.0	-	0.0	-	0.0	-	-0.0015	(0.0004)
κ_{21}	0.0911	(0.0316)	0.2044	(0.1256)	-0.0808	(0.0172)	0.0050	(0.0146)
κ_{22}	0.1622	(0.0201)	1.3502	(0.0877)	0.4404	(0.0081)	0.0841	(0.0053)
κ_{23}	0.0		-3.9500	(0.5703)	0.0	-	0.0001	(0.0002)
κ_{31}	0.4306	(0.1092)	-0.2025	(0.0767)	0.1030	(0.2555)	-0.0004	(0.0002)
κ_{32}	0.3111	(0.0322)	2.0133	(0.1211)	0.2531	(1.1114)	0.0210	(0.0037)
κ_{33}	0.0949	(0.0337)	0.0728	(0.0255)	0.2467	(0.0959)	0.7813	(0.4767)
θ_1	0.0	-	0.1510	(0.0667)	0.0928	(0.0316)	0.0075	(0.0021)
θ_2	0.0	-	0.0	-	0.1334	(0.0677)	0.5009	(0.2088)
θ_3	0.0	-	0.0	-	0.0	-	0.0062	(0.0029)

α_1	1.0	-	0.0	-	0.0	-	0.0	-
α_2	1.0	-	1.0	-	0.0	-	0.0	-
α_3	1.0	-	1.0	-	1.0	-	0.0	-
β_{11}	0.0	-	1.0	-	1.0	-	1.0	-
β_{12}	0.0	-	0.0	-	0.0	-	0.0	-
β_{13}	0.0	-	0.0	-	0.0	-	0.0	-
β_{21}	0.0	-	0.0505	(0.0028)	0.0	-	0.0	-
β_{22}	0.0	-	0.0	-	1.0	-	1.0	-
β_{23}	0.0	-	0.0	-	0.0	-	0.0	-
β_{31}	0.0	-	0.3130	(0.0114)	0.1015	(0.5467)	0.0	-
β_{32}	0.0	-	0.0	-	0.0202	(0.0041)	0.0	-
β_{33}	0.0	-	0.0	-	1.0	-	1.0	-
λ_0/λ_0^Q	0.0520/0.1008	(0.0288)/(0.0987)	0.0994/0.1712	(0.0088)/(0.1032)	0.0307/0.1106	(0.0633)/(0.1266)	0.0035/0.0206	(0.0055)/(0.0063)
$\lambda_{11}/\lambda_{11}^Q$	0.0402/0.1175	(0.0282)/(0.0367)	0.1167/0.3422	(0.0222)/(0.0268)	0.0550/0.3214	(0.0012)/(0.1971)	0.0033/0.0505	(0.0011)/(0.0279)
$\lambda_{12}/\lambda_{12}^Q$	0.0523/0.1317	(0.0075)/(0.0333)	0.1233/0.4244	(0.0167)/(0.1521)	0.1005/0.4971	(0.1114)/(0.5772)	0.0020/0.6002	(0.0016)/(0.1024)
$\lambda_{13}/\lambda_{13}^Q$	0.0624/0.0707	(0.0015)/(0.0409)	0.0955/0.2030	(0.0056)/(0.0304)	0.0416/0.2248	(0.0029)/(0.1270)	0.0070/0.0404	(0.0048)/(0.0302)
μ_1/μ_1^Q	-0.0212/-0.0277	(0.0586)/(0.0155)	-0.0444/-0.1547	(0.0628)/(0.0651)	-0.0333/-0.1001	(0.9372)/(0.0384)	-0.0650/-0.1103	(0.0174)/(0.0411)

μ_2/μ_2^Q	-0.0118/-0.0444	(0.0018)/(0.0260)	-0.0320/-0.1210	(0.0284)/(0.0405)	-0.0220/-0.0443	(0.0105)/(0.0104)	-0.0102/-0.0567	(0.0052)/(0.0133)
μ_3/μ_3^Q	-0.0219/-0.1409	(0.0055)/(0.0812)	-0.0334/-0.1049	(0.0033)/(0.0426)	-0.0116/-0.0544	(0.0052)/(0.0143)	-0.0230/-0.0530	(0.0014)/(0.0465)
σ_1^2	0.0404	(0.0222)	0.4005	(0.0732)	0.0228	(0.0227)	0.0270	(0.0158)
σ_2^2	0.0354	(0.0067)	0.1450	(0.1143)	0.0116	(0.0067)	0.0177	(0.0067)
σ_3^2	0.0720	(0.0246)	0.1152	(0.0228)	0.0188	(0.0052)	0.0232	(0.0066)
η_{11}	-0.2667	(0.0198)	-0.0740	(0.0149)	-0.0433	(0.0979)	-0.0808	(0.0621)
η_{12}	-0.1373	(0.0752)	-0.1004	(0.0122)	-0.0515	(0.0124)	-0.1111	(0.0273)
η_{13}	-0.3141	(0.0204)	-0.1746	(0.0251)	-0.0339	(0.0081)	-0.4850	(0.0669)
δ_0	-0.0337	(0.0035)	0.0537	(0.0075)	0.0170	(0.0067)	0.0708	(0.0574)
δ_{11}	0.0620	(0.0111)	0.0631	(0.0111)	0.0702	(0.0366)	0.0904	(0.0572)
δ_{12}	0.1203	(0.0433)	0.0967	(0.0414)	0.2511	(0.1480)	0.1040	(0.0910)
δ_{13}	0.0555	(0.0162)	-0.1276	(0.8707)	-0.0777	(0.0113)	-0.0303	(0.0045)
$\eta_{2,11}$	-0.0101	(0.0206)	0.0	-	0.0	-	0.0	-
$\eta_{2,12}$	-0.0200	(0.0048)	0.0	-	0.0	-	0.0	-
$\eta_{2,13}$	-0.0360	(0.0022)	0.0	-	0.0	-	0.0	-
$\eta_{2,21}$	-0.0410	(0.0317)	-0.3589	(0.8787)	0.0	-	0.0	-
$\eta_{2,22}$	-0.0705	(0.1996)	-0.7088	(0.5622)	0.0	-	0.0	-

$\eta_{2,23}$	-0.0606	(0.0098)	-0.1860	(0.0365)	0.0	-	0.0	-
$\eta_{2,31}$	-0.0702	(0.1095)	-0.3200	(0.1853)	-0.0360	(0.2545)	0.0	-
$\eta_{2,32}$	-0.0900	(0.0333)	-0.0450	(0.2602)	-0.0880	(0.1852)	0.0	-
$\eta_{2,33}$	-0.0237	(0.0031)	-0.2179	(0.1684)	-0.0433	(0.9664)	0.0	-

This table reports parameter estimates for three-factor essentially AJD models using LIBOR-Swap rates with maturities of 6-month, 2-year, 10-year, 3-year, 5-year, and 7-year and sampled daily from August 13, 1990 to December 31, 2008. The rates with the first three maturities are assumed to be observed without error and the remaining three observed with error. Estimation follows the infinitesimal operator based procedure described in Section 3.3.2. The specifications of the models, $\text{EAJD}_0(3)$, $\text{EAJD}_1(3)$, $\text{EAJD}_2(3)$ and $\text{EAJD}_3(3)$, have the sport rate as $r_t = \delta_0+\delta_{11}X_{1t}+\delta_{12}X_{2t}+\delta_{13}X_{3t}$, risk-neutral dynamics in (3.21), (3.23), (3.25) and (3.27), and physical dynamics in (3.120), (3.22), (3.21) and (3.26) for X_t. Reported in the parentheses of the ”SE” column are the standard errors of the coefficients estimates. The blanks in the ”SE” columns refer to those parameters pre-specified by the model structures.

-ression") and in Dai and Singleton (2002) using U.S. Treasury yields (both the $EA_0(3)$ and $EA_1(3)$ models match the unity line in the "risk-adjusted yield regression").

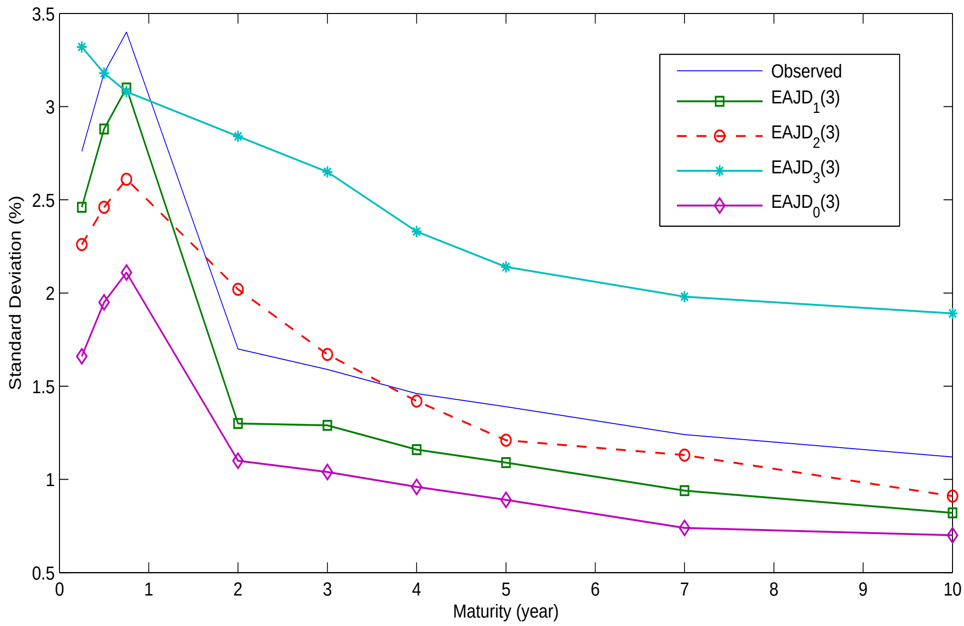
Another observation similar to what we have seen with respect to AD models in Section 3.4.3 is that, for all four subclasses, Figure 3.6.B shows a gap between the risk-adjusted coefficient ϕ_τ^R and unity for the LIBOR-Swap rates with maturities of under two years. For example, the risk-adjusted coefficient ϕ_τ^R is only around 0.92 when $\tau = 9$ months. As said in Section 3.4.3, I shall investigate this issue for the short end of the term structure movements in Section 3.4.5.

Time-Varying Conditional Volatilities. For the volatility persistence, Table 3.5 reports parameter estimates of the GARCH(1,1) model for both the historical and simulated 5-year yield. Similar to the model $EA_0(3)$, the model $EAJD_0(3)$ still exhibits little volatility persistence. Hence, the introduction of jumps into state variables does not help the $EAJD_0(3)$ model as the volatility functions are still assumed to be constants. In contrast, the implied GARCH(1,1) estimates form the other three models in the three-factor "essentially" AJD class, $EAJD_1(3)$, $EAJD_2(3)$ and $EAJD_3(3)$, match those in the sample quite closely. In fact, we can see that these three AJD models perform better than their corresponding AD models in matching the degree of volatility persistance, with higher β estimates. This confirms that fact that jumps in state variables have impacts on volatility dynamics once the state variables enter the specification of volatilities.

Figure 3.7 plots sample variances computed using simulated time series of LIBOR-Swap rates from the models evaluated at their estimated parameter values. In terms of the shape, we can see that except for the $EAJD_3(3)$ model,

Figure 3.7: **The AJD Model-Implied Term Structure of Volatilities**

This figure plots the term structure of the sample variances, computed using simulated time series of daily changes in the logarithms of LIBOR-Swap rates from the four three-factor essentially AJD models ($EAJD_0(3)$, $EAJD_1(3)$, $EAJD_2(3)$, and $EAJD_3(3)$) evaluated at their estimated parameter values. The "observed" line refers to those computed by the historical rates.



the other three models ($EAJD_1(3)$, $EAJD_2(3)$, and $EAJD_3(3)$) all have a hump at around the one-year maturity. One important difference to note is that, by comparing Figure 3.7 with Figure 3.5, we can see that all four three-factor essentially AJD term structure models have larger magnitudes of simulated variances than do their AD counterparts without jumps. The magnitude of sample variances of the $EAJD_0(3)$ model is still too small to match the historical level. But the sizes of simulated sample variances from both $EAJD_1(3)$ and $EAJD_2(3)$ are pretty close to the historical level; the $EAJD_1(3)$ model performs better at the longer maturities and the $\mathrm{EAJD}_2(3)$ model performs better at the short end.

In summary, the empirical finding here is that, unlike the three-factor essentially AD models, all four models in the three-factor essentially AJD class match time-varying risk premiums closely. Furthermore, both the $\mathrm{EAJD}_1(3)$ and $\mathrm{EAJD}_2(3)$ models simultaneously match time variations in the risk premium and conditional volatility closely. Although the $\mathrm{EAJD}_0(3)$ model with the most flexible specification for the market prices of diffusive risks still performs the best in matching time-varying risk premiums, the $\mathrm{EAJD}_3(3)$ model with the most flexible time-varying conditional volatility specification is also able to match time variations in the risk premium reasonably closely. Results obtained for AJD models using the LIBOR-Swap yields suggest that, contrary to the empirical evidence for AD models in Section 3.4.3 and Dai and Singleton (2002), there does not exist any (serious) tension between matching first-order and second-order moments of the interest rates movements. I shall provide theoretical support for this empirical success in the following by analyzing the relationship between the market prices of risks and time-varying conditional volatility of the risk factors in AJD models.

As pointed out in Section 3.4.3, the tension in AD models that arises when matching the time variations in the risk premium and conditional volatility simultaneously is due to the specification of the market prices of risk $\Lambda_t = \sqrt{S_t}\eta_1 + \sqrt{S_t^-}\eta_2 X_t$, which leads to a trade-off between more flexible market prices of risks with a smaller m, for which more elements of risk premiums can depend on X_t directly and switch signs, and more flexible specifications of time-varying conditional volatilities with a bigger m, for which more risk factors drive the conditional volatility. For AJD models, market prices are introduced for both the jump size and jump timing risks as:

$$\Lambda_t^J = \lambda_0 \mu - \lambda_0^Q \mu^Q + \left[(\lambda_1)' \mu - \left(\lambda_1^Q \right)' \mu^Q \right] X_t \tag{3.55}$$

from (3.18). By (3.19), the total market prices of risks in AJD models are specified as

$$\Lambda_t^{AJD} = \sqrt{S_t} \eta_1 + \sqrt{S_t^-} \eta_2 X_t + \Lambda_t^J, \tag{3.56}$$

where the first two terms are market prices of diffusive risks and the last term denotes the market prices of jump risks in (3.55). Though the tension still exists between flexible market prices of diffusive risks and time-varying conditional volatilities, the specification of market prices of risks Λ_t^J contains a term which is able to switch signs at no cost in terms of time-varying conditional volatilities, i.e., $\left[(\lambda_1)' \mu - \left(\lambda_1^Q \right)' \mu^Q \right] X_t$ due to its direct dependence on X_t and introduced only by jumps in risk factors. The key that allows AJD term structure models to match time variations in both the risk premium and conditional volatility simultaneously is this jump risk premium Λ_t^J, which generalizes the essentially affine market prices of risks Λ_t without imposing any single restriction on the time-varying conditional volatility.

In three-factor essentially AD models, the only model which has three elements of the market prices of risks that are able to switch signs is the $EA_0(3)$ model with no factors entering the conditional volatility. In contrast, all four models in the three-factor essentially AJD class, including the $EAJD_3(3)$ model with three factors driving time variations in the conditional volatility, have this feature (three sign-switching elements in the market prices of risks) since Λ_t^{AJD} depends on X_t directly in all three components. Therefore, the jump risk premium leads to more flexible time-varying market prices of risk without restricting the time-varying conditional volatility. Consequently, it does break up the tension in affine term structure models between a more flexible risk premium specification that is important to capturing time-varying risk premiums and a more flexible time-varying conditional volatility structure that is critical to matching time-varying conditional volatility. This explains the empirical evidence shown in the last section that both the $\text{EAJD}_1(3)$ and $\text{EAJD}_2(3)$ models simultaneously match time variations in the risk premium and conditional volatility closely and even the $\text{EAJD}_3(3)$ model, with the most flexible time-varying conditional volatility, can match time-varying risk premiums reasonably closely.

Note as well that the introduction of jump risk premiums does not affect the conditional correlation structure. Hence, combined with the last observation made in Section 3.4.3.2 that there is also a trade-off between conditional correlation and conditional volatility, the rank order $\text{EAJD}_0(3)$ ¿ $\text{EAJD}_1(3)$¿ $\text{EAJD}_2(3)$¿ $\text{EAJD}_3(3)$ in terms of their performance in capturing time-varying risk premiums in fact confirms the finding in Dai and Singleton (2002) that a flexible conditional correlation structure also helps to fit time-variations in the risk premium.

In trying to resolve the tension between matching the mean and volatility of interest rates documented in Dai and Singleton (2002) and Duffee (2002), Duarte (2004) proposes the "Semi affine square-root" (SAS-R) model, which extends the completely and essentially AD models in Dai and Singleton (2000) and Duffee (2002) by the following flexible specification for the market prices of risk:

$$\Lambda_t^{SAS-R} = \sqrt{S_t}\eta_1 + \sqrt{S_t^-}\eta_2 X_t + \Sigma^{-1}\eta_0, \tag{3.57}$$

where η_0 is a $d\times 1$ vector. The new risk premium term $\Sigma^{-1}\eta_0$, in addition to the essentially affine risk premium $\sqrt{S_t}\eta_1 + \sqrt{S_t^-}\eta_2 X_t$, offers additional sign-switching flexibility at no expense in terms of limiting the volatility dynamics. The empirical finding in Duarte (2004) is that the SAS-R specification of the market prices of risks, though producing improvement in capturing time variations in the risk premium, is not sufficient to solve the mean-volatility tension.

Comparing (3.57) with (3.56), we can see that Λ_t^{AJD} and Λ_t^{SAS-R} share the same idea of generalizing the essentially affine market prices of risks to allow sign changes in all elements of the market prices of risks. The difference is that the jump risk premium is introduced naturally by a well-documented stylized fact, that is, jumps in interest rates, besides the role of capturing time-varying risk premiums. Moreover, the jump risk premium depends on X_t directly and is more flexible than Duarte's (2004) SAS-R specification through Σ^{-1}. The latter may impose a restriction on the conditional correlation structure, which explains the empirical finding of Duarte (2004) that the SAS-R model is not sufficient to solve the mean-volatility tension in light of Dai and Singleton's (2002) finding that a flexible conditional correlation is also necessary to capture time-varying risk premiums.

3.4.5 The Short End of LIBOR-Swap Yields

As documented in Sections 3.4.3 and 3.4.4, both three-factor essentially AD and AJD term structure models fail to match time-varying conditional risk premiums at the short end of the LIBOR-Swap curve. Using U.S. Treasury yields, Dai and Singleton (2002) show that the mismatch at the short end can be rectified by the addition of a fourth short-end factor.[17] In this section, I check whether adding one more factor to three-factor essentially AJD models can cor-

[17] Piazzesi (2005) shows the importance of the fourth factor, related to announcements of the Federal Open Market Committee, in capturing the short end of the term structure dynamics in terms of explaining the snake shape of the volatility term structure.

rect the mismatch of the unity line in the "risk-adjusted yield regression" at the short-end of the term structure dynamics. I choose to rectify the mismatch of the $EAJD_0(3)$ model and consider the $EAJD_0(4)$ model as the rescue since the former is shown to have the best match of time-varying risk premiums at the longer maturities. I estimate the $EAJD_0(4)$ model using 6-month, 2-year, 5-year and 10-year rates as observed without error and 9-month, 3-year, 4-year and 7-year rates as those observed with error.[18] Then the sample counterparts $\phi^R_{\tau T}$ of the population coefficients ϕ^R_τ in (3.54) are computed and reported in Figure 3.8 along with the historical projection coefficients from Table 3.2. We can see that as is true in Dai and Singleton (2002) for AD term structure models, the $EAJD_0(4)$ has a nearly perfect match with the unity line (see Dai and Singleton (2002) for details on why the omission of the fourth factor can potentially lead to the failure of matching the unity line in the "risk-adjusted yield regression" at the short end). This shows that a fourth factor is important if we are to capture the short end of the term structure dynamics (See Longstaff et al. (2001) for further discussion).

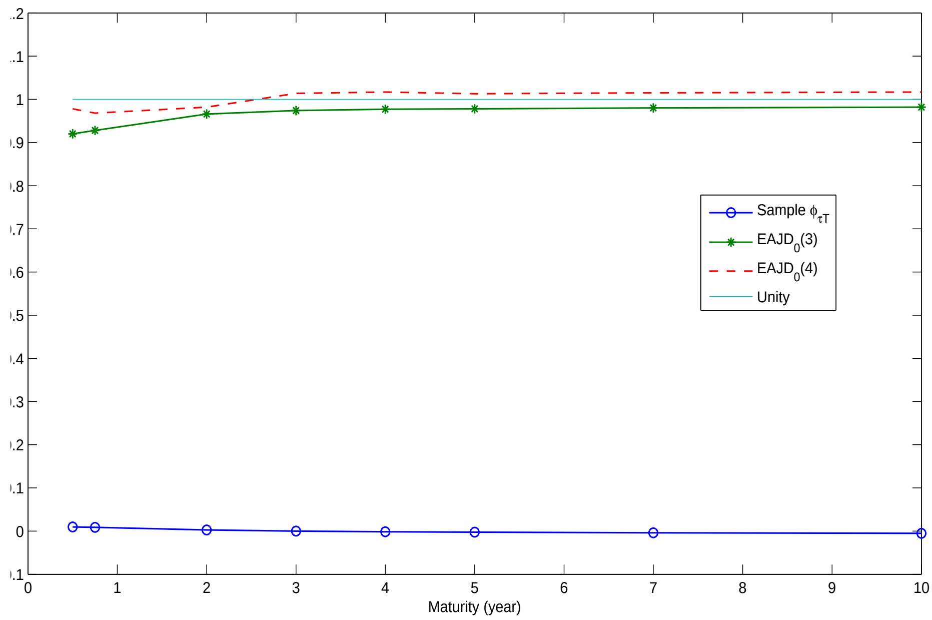
3.5 Jump Risk Premiums

In this section, I shall first discuss the details of the estimation and then present features of jumps in the term structure dynamics of the LIBOR-Swap yields. As noted in Section 3.4.2, all three-factor essentially AD and AJD term structure models are estimated. Following Ait-Sahalia and Kimmel (2010), I use a number of yields that is equal to twice the number of factors in the model for the esti-

[18] The model parameter estimates are not reported here, to save space. They are available upon request from the author.

Figure 3.8: **AJD Model-Implied Patterns for the "Risk-Adjusted Yield Regression"**

This figure plots implied estimates of the slope coefficients ϕ^R_τ from the "risk-adjusted yield regression" of $y^{(\tau-1)}_{t+1} - y^\tau_t + e^\tau_t / (\tau - 1)$ onto $(y^\tau_t - r_t) / (\tau - 1)$, where e^τ_t denotes the time-t expected excess holding period return on the zero-coupon bond with τ periods to mature, for the $EAJD_0(3)$ and $EAJD_0(4)$ models. "Sample $\phi_{\tau T}$" displays estimated coefficients reported in Table 2 by the sample "yield regression" using daily LIBOR-Swap rates from August 13, 1990 to December 31, 2008.



mation, the daily ($\Delta = 1/252$) LIBOR-Swap yields with maturities of 6 months, 2 years, 3 years, 5 years, 7 years and 10 years. The six yields I choose have the same maturity structures as those in Dai and Singleton (2002) using monthly U.S. Treasury yields. As discussed in Section 3.3.2, choices need be made as to which yields are assumed to be observed with and without errors for the estimation. I assume that the yields of 6-month, 2-year, and 10-year maturities are measured without error while those of 3-year, 5-year, and 7-year maturities are observed with error. As argued by Duffee (2002) and Dai and Singleton (2002), this choice can span as much of the term structure as possible without treating the 3-month LIBOR rate (it may involve idiosyncratic dynamics) as observed without error.

The parameter estimates with standard errors for three-factor essentially AD models are presented in Table 3.4 while those for AJD models are presented in Table 3.6. Choices pertaining to the trimming parameter ς, the kernel function $\mathcal{K}(\cdot)$, and the bandwidth b_n for computing the estimators follow Chapter 2. Tables 3.4 and 3.6 contain some blank entries since some parameters are constrained as constants by the admissibility conditions (see Section 3.2.3 for details). For these estimation results, we focus on the significance of jump parameters and discuss specifically the jump risk premiums. First from Table 3.6, it can be observed that most of the parameters in jump specifications are significantly different from zero, implying the importance of modeling jumps in term structure models. In particular, the significance of jump intensity parameters $\lambda' s$ show that the jump arrivals are indeed state-dependent in the term structure dynamics, implying a certain degree of predictability of current market conditions for the frequency of future large changes in bond yields. Second, overwhelming evidence of negative jumps are found in the state-variable dynamics under both physical and risk-neutral measures. Similar to what obtains in Jarrow, Li and Zhao (2007), this may reflect investors' fears of a market crash such as the one that occurred in 1987.

Finally, the risk premiums are estimated to be positive. For example, the coefficient $(\mu_1 - \mu_1^Q, \mu_2 - \mu_2^Q, \mu_3 - \mu_3^Q)$ associated with the premiums for jump-size uncertainty is estimated to be (6.68%, 2.23%, 4.28%) while the coefficient $(\lambda_0 - \lambda_0^Q, \lambda_{11} - \lambda_{11}^Q, \lambda_{12} - \lambda_{12}^Q, \lambda_{13} - \lambda_{13}^Q)$ associated with the premia for jump-arrival uncertainty is estimated to be (−7.99%, −26.64%, −39.66%, −18.32%) for the $\text{EAJD}_2(3)$ model. Combined, they imply a positive jump risk premium of $\lambda_0\mu - \lambda_0^Q\mu^Q + \left[(\lambda_1)'\mu - \left(\lambda_1^Q\right)'\mu^Q\right]X_t$ =(1.00%, 0.42%, 0.57%)+6.19% $\cdot X_t$. Moreover, the large discrepancy between jump intensities under the physical and risk-neutral measures are similar to that for jump sizes in Jarrow, Li and Zhao (2007) for LIBOR rates and in Pan (2002) for index options. This could be the result of a huge jump risk premium. For the diffusive risk, it can be observed that $\eta_1 = (-4.33\%, -5.15\%, -3.39\%)$ for the part proportional to $\sqrt{S_t}$ and the nonzero elements of η_2 are $(\eta_{2,31}, \eta_{2,32}, \eta_{2,33}) = (-3.60\%, -8.80\%, -4.33\%)$ for the part that can depend on X_t directly. This implies a negative volatility risk premium consistent with Dai and Singleton (2000) using weekly LIBOR-Swap rates and with Carr and Wu (2009) using the S&P 500 index.

Note that these observations actually answer the first two questions raised by Johannes and Polson (2009): "Do multiple factors jump, or is it only the short rate? Does the market price diffusive and jump risks differently in the term structure?". In particular, the significance of the jump parameters in the state variable dynamics implies that multiple risk factors do jump. Moreover, the market prices diffusive and jumps risks differently: jump risk premiums are positive while the volatility risk premium is negative.

3.6 Conclusion

In this paper, I develop a multivariate AJD term structure model to solve the two empirical challenges associated with AD models. First, features of jumps are documented for the term structure dynamics of the LIBOR-Swap curve. In particular, I find that jumps are state-dependent and negative. The risk premium is positive for jump size risk and negative for jump timing risk while the total jump risk premium is positive. Second, two models in the three-factor AJD class simultaneously match time variations in both the risk premium and conditional volatility of the LIBOR-Swap yields. The key to this empirical success is the jump risk premium, which leads to flexible time-varying market prices of risk without restricting time variations in the conditional volatility. This in fact answers the question raised by Duffee (2002): "It remains to be seen whether an essentially affine model can be constructed that reproduces the time variation observed in both the conditional variances of yields and expected returns to bonds."

Because it is difficult for AD models to capture time-varying risk premiums and conditional volatilities simultaneously, many studies have considered non-affine term structure models for the term structure movements, including, for example, the semi-affine model in Duarte (2004), the quadratic-Gaussian models in Ahn et al. (2002) and Leippold and Wu (2002), and the regime-shift model in Bansal and Zhou (2002). Contrary to commonly held beliefs in the literature, however, the empirical finding in this paper shows that affine models are able to capture time-varying risk premiums and conditional volatilities simultaneously when jumps are introduced into the models. In contrast, term structure models like quadratic-Gaussian models do not allow jumps and cannot capture this well-documented stylized fact about interest rates (Chen, Bayraktar and Poor, 2005).

For the two models ($EAJD_1(3)$ and $EAJD_2(3)$) that can match time variations in the risk premium and conditional volatility simultaneously, differences still exist in their performances. For example, Section 5.4.1 documents that the $EAJD_1(3)$ model performs a bit better at matching time-varying risk premiums than the $EAJD_2(3)$ model does. In terms of matching time-varying conditional volatilities, however, the $EAJD_2(3)$ model is better at the longer maturities although the $EAJD_1(3)$ model performs better at the short end of LIBOR-Swap yields. Hence it is not absolutely clear which model we should choose for modeling the term structure dynamics of the LIBOR-Swap curve. More highly detailed comparisons, ideally based on some econometric procedures which could draw conclusions with statistical significance, should prove valuable. In a follow-up work, Li and Chapter 2 take up this challenge, proposing a sequence of rigorous model selection tests and conducting comprehensive comparisons of three-factor AD and AJD term structure models.

Explaining the violation of the "expectation hypothesis" for interest rates is actually equivalent to explaining time-varying bond returns (Duffee, 2002), which is further related to the bond return predictability (Fama and Bliss, 1987; Cochrane and Piazzesi, 2005). It would be interesting to see if AJD term structure models can replicate the documented predictability evidence; see Dai, Singleton and Yang (2004) and Bansal, Tauchen, and Zhou (2004) for similar studies using AD models and models with regime shifts. Another interesting future extension is to investigate how AJD term structure models perform in pricing fixed-income derivatives, such as caps, swaptions, and so on. In light of the empirical success of AJD models to capture time variations in both the risk premium and conditional volatility, it is interesting to study whether the "unspanned stochastic volatility" is still true with the conclusion that risk factors in the prices of caps and swaptions are not spanned by the underlying LIBOR-Swap rates(Heidari and Wu, 2003; Collin-Dufresne and Goldstein, 2002; Li and Zhao, 2006; Joslin, 2007; Collin-Dufresne, Goldstein and Jones, 2009; Andersen and Benzoni , 2010; Bikbov and Chernov, 2010).

CHAPTER 4

A MARTINGALE APPROACH FOR TESTING DIFFUSION MODELS BASED ON INFINITESIMAL OPERATOR

4.1 Infinitesimal Operator Based Martingale Characterization

Consider a multivariate diffusion model defined by the following stochastic differential eqnarray (SDE):

$$dX_t = b^0(X_t)dt + \sigma^0(X_t)dW_t \tag{4.1}$$

where W_t is a $d \times 1$ standard Brownian motion in $\mathbb{R}^d$, $b : E \subset \mathbb{R}^d \to \mathbb{R}^d$ is a drift function (i.e., instantaneous conditional mean) and $\sigma : E \to \mathbb{R}^{d\times d}$ is a diffusion function (i.e., instantaneous conditional standard deviation). We will call (4.1) a SDE-diffusion process.

What we are interested in is to test the parametric form of a SDE-diffusion, i.e.,

$$\begin{aligned} b^0 &\in \mathcal{M}_b \triangleq \{b(\cdot,\theta), \theta \in \Theta\} \\ \sigma^0 &\in \mathcal{M}_\sigma \triangleq \{\sigma(\cdot,\theta), \theta \in \Theta\} \end{aligned} \tag{4.2}$$

where Θ is a finite-dimensional parameter space. We say that the model $\{\mathcal{M}_b, \mathcal{M}_\sigma\}$ is correctly specified for (4.1) if

$$H_0 : P[b(X_t,\theta_0) = b^0(X_t), \sigma(X_t,\theta_0) = \sigma^0(X_t)] = 1 \tag{4.3}$$

for some $\theta_0 \in \Theta$. The alternative hypothesis is that there exists no parameter value $\theta \in \Theta$ such that $b(\cdot,\theta)$ and $\sigma(\cdot,\theta)$ coincide with $b^0(\cdot)$ and $\sigma^0(\cdot)$ simultaneously:

$$H_A : P[b(X_t,\theta) = b^0(X_t), \sigma(X_t,\theta) = \sigma^0(X_t)] < 1 \tag{4.4}$$

for all $\theta \in \Theta$.

Since in this study I am relying on a characterization of a continuous-time Markov process alternative to transition density, i.e., the infinitesimal operator, and in finance a diffusion process is usually specified as a SDE-diffusion, I will discuss first some related mathematical concepts and clarify their relationship. By Rogers and Williams(2000, Ch III.1), a continuous time Markov process is defined as follows:

Definition 4.1.1: A Markov process $X = (\Omega, \{\mathfrak{F}_t\}, \{X_t\}, \{P_t\}, \{P^x, x \in E\})_{t \geqq 0}$ with state space (E, ε) is an E-valued stochastic process adapted to the sequence of σ-algebras $\{\mathfrak{F}_t\}$ such that

for $0 \leqq s \leqq t$ and $x \in E$, $E^x[f(X_{s+t})|\mathfrak{F}_s] = (P_t f)(X_s)$, P^x-$a.s.$

where $\{P_t\}$ is a transition function on (E, ε), i.e., a family of kernels $P_t : E \times \varepsilon \rightarrow [0, 1]$ such that

(i): for $t \geqq 0$ and $x \in E$, $P_t(x, \cdot)$ is a measure on ε with $P_t(x, E) \leqq 1$

(ii): for $t \geqq 0$ and $\Gamma \in \varepsilon$, $P_t(\cdot, \Gamma)$ is ε-measurable

(iii): for $s, t \geqq 0$, $x \in E$ and $\Gamma \in \varepsilon$,

$$P_{t+s}(x, \Gamma) = \int_E P_s(x, dy) P_t(y, \Gamma) \tag{4.5}$$

In this definition, the Markov property is characterized by the transition function (or transition density when the density of transition function exists) and (4.5) is the so-called Chapman-Kolmogrov eqnarray. An alternative and equivalent characterization is the induced family $\{P_t\}$ which is a set of positive bounded operators with norm less than or equal to 1 on $b\varepsilon$(bounded and

ε-measurable functions) and which is defined by:

$$P_t f(x) \equiv (P_t f)(x) = \int_E P_t(x, dy) f(y) \tag{4.6}$$

In this case, the markov property is expressed as the following semi-group property equivalent to the Chapman-Kolmogrov eqnarray:

$$P_s P_t = P_{s+t}, \tag{4.7}$$

for any $s, t \geqq 0$.

Both transition function and the semi-group of operators characterize the Markov process and interact with the sample-path property of the process. However, since the general Markov process consists of too many processes and is too broad, we choose to focus on the more interesting subclass, Feller process. By Rogers and Williams(2000, Ch III.6), Feller process is defined as follows :

Definition 4.1.2: The transition function $\{P_t\}_{t \geqq 0}$ of a Markov process is called a Feller transition function if

(i): $P_t C_0 \subset C_0$ for all $t \geqq 0$

(ii): for any $f \in C_0$ and $x \in E$, $P_t f(x) \to f(x)$ as $t \downarrow 0$, where $C_0 = C_0(E)$ is the space of real-valued, continuous functions on E which vanish at infinity and C_0 is endowed with the sup-norm.

Feller process has good path properties[1] and is also general enough to contain most processes we are interested in, for example, Feller diffusion which will be defined below and has been extensively used in finance, and Levy process including Poisson process and Compound process which has received more and

[1]By Rogers and Williams(2000, Ch III.7-9), the canonical Feller process always admits a Cadlag(the path of the process is right continuous and has left limits) modification and satisfies the strong Markov property

more attention in finance recently(see Schoutens, 2003). For Feller processes, we will consider another characterization, the infinitesimal operator, other than the transition function and semi-group of operators introduced above which are for the general Markov process.

Definition 4.1.3: A function $f \in C_0$ is said to belong to the domain $D(\mathcal{A})$ of the infinitesimal operator of a Feller process X if the limit

$$\mathcal{A}f = \lim_{t\downarrow 0} \frac{P_t f - f}{t} \tag{4.8}$$

exists in C_0. The linear transformation $\mathcal{A} : D(\mathcal{A}) \to C_0$ is called the infinitesimal operator of the process.

Immediately from Definition 4.1.3, we see that for $f \in D(\mathcal{A})$, it holds P-$a.s.$ that

$$E\left(\frac{f(X_{t+h}) - f(X_t)}{h}|\mathfrak{F}_t\right) = \mathcal{A}f(X_t) + o(h) \tag{4.9}$$

as $h \downarrow 0$. In this sense, the infinitesimal operator indeed describes the movement of the process in an infinitesimally small amount of time. Therefore, intuitively the infinitesimal operator characterizes the whole dynamics of a Feller process because the time is continuous here[2].

So far we have had Feller process for which three complete characterization of the dynamics are available: transition function(or transition density), semi-group of operators and infinitesimal operator. The most important Feller process in continuous time finance is the diffusion process. By Rogers and Williams(2000, ChIII.13),

[2]Rigorously, it can be proved that the infinitesimal operator is equivalent to the semi-group of operators in characterizing a Feller process(see the Hill-Yoshida theorem in Dynkin(1965)). Therefore, infinitesimal operator does determine the whole dynamics of the process.

Definition 4.1.4: A Feller process with state space $E \subset \mathbb{R}^d$ is called a Feller diffusion if it has continuous sample paths and the domain of its infinitesimal operator contains the function space $C_c^\infty(int(E))$ which is the space of infinitely differentiable functions with compact support contained in the interior of the state space E.

We can see that the Feller diffusion is defined through the combination of the sample path properties and the restrictions imposed on the infinitesimal operator. A very convenient property of Feller diffusion is that its infinitesimal operator has an explicit form. According to Kallenberg(2002, Thm 19.24) and Rogers and Williams (2000, Vol1, Thm III.13.3 and Vol2, Ch V.2), for a Feller diffusion $\{X_t\}$, there exist some functions $a_{i,j}$ and $b_i \in C(\mathbb{R}^d)$ for $i, j = 1, \cdots, d$ where $\left(a_{i,j}\right)_{i,j=1}^d$ forms a symmetric nonnegative definite matrix such that the infinitesimal operator is

$$\mathcal{A}f(x) = \sum_{i=1}^{d} b_i(x) f_i'(x) + \frac{1}{2} \sum_{i,j=1}^{d} a_{i,j}(x) f_{i,j}''(x) \tag{4.10}$$

for $f \in D(\mathcal{A})$ and $x \in \mathbb{R}^d$.

Now we have arrived at the Feller diffusion and its infinitesimal operator which has a closed form. Then what is the relationship between Feller diffusion and SDE-diffusion (4.1)? By Rogers and Williams(2000, ChV.2 and V.22), under some regularity conditions, they are equivalent. That is, for a Feller diffusion as in Definition4, there is a corresponding SDE-diffusion and also a SDE-diffusion like (4.1) is a Feller diffusion, where the function $b(\cdot)$ are the same and $a = \sigma\sigma^T$, i.e., $a_{ij}(x) = \sum_{k=1}^d \sigma_{i,k}(x)\sigma_{j,k}(x)$. Therefore, the SDE-diffusion which has been analyzed extensively in continuous-time finance and which belongs to the class of Feller process also has (4.10) as the closed-form infinitesimal operator.

To illustrate the relationship between infinitesimal operator and drift and diffusion terms, let's consider the univariate diffusion defined as $dX_t = b(X_t)dt + \sigma(X_t)dW_t$ with W_t a 1-dimensional standard Brownian motion in $\mathbb{R}$, $b : E \subset \mathbb{R} \to \mathbb{R}$ a drift function and $\sigma : E \to \mathbb{R}$ a diffusion function. Then by (4.10) and the discussion above, the infinitesimal operator for this univariate diffusion is

$$\mathcal{A}f(x) = b(x)f'(x) + \frac{1}{2}\sigma^2(x)f''(x) \tag{4.11}$$

Clearly the first term involving the first derivative of function $f(\cdot)$ is related to the dynamics of drift and the second term involving the second derivative of function $f(\cdot)$ to the dynamics of diffusion function. This is consistent with the intuition that drift describes the dynamics of mean and the diffusion describes that of variance of the process(see Nelson 1990 for more discussion which proves that the diffusion process is the approximation of an ARCH process). However, we have to point out that it is not absolutely right to simply think of drift and diffusion terms as the continuous time counterparts of conditional mean and variance respectively.. Consider the infinitesimal changes of this univariate diffusion process. By (4.9) and (4.11), for any $f \in D(\mathcal{A})$, it holds P-$a.s.$ that

$$E\left(\frac{f(X_{t+h}) - f(X_t)}{h}|\mathfrak{F}_t\right) = b(X_t)f'(X_t) + \frac{1}{2}\sigma^2(X_t)f''(X_t) + o(h), \tag{4.12}$$

as $h \downarrow 0$. Therefore, the dynamics of $\{X_t\}$ are characterized completely by the drift and diffusion coefficients, including the conditional probability law. But in discrete time series models, the mean and variance solely cannot determine the complete conditional probability law unless it is Gaussian. In fact, the conditional mean of the process $\{X_t\}$, $E\left[X_{t+h}|X_t\right]$ for a fixed $h > 0$ is in general a function of both the drift $b(\cdot)$ and diffusion $\sigma(\cdot)$ instead of the drift solely(see Ait-Sahalia 1996a).

From the discussions above, for a SDE-diffusion which is also a Feller diffusion, there are at least two characterizations we can use to identify the whole dynamics of the process: the transition function and the the closed-form infinitesimal operator which is also the generator of the third characterization, semi-group of operators. The former, also well known as transition density when the density of transition function exists, has been the primary tool to analyze the diffusion process, not only in estimation(see Ait-Sahalia 2002b) but also in the construction of specification tests (see Hong and Li 2005, Ait-Sahalia, Fan and Peng 2008). However, as we discussed in Section1, specification of drift and diffusion terms rarely give a closed-form transition density. In contrast, (4.10) and (4.11) tell us that the infinitesimal operator does have a direct and explicit expression and this nice property makes it a convenient tool for analyzing the diffusion process. It has already been used in identification and estimation problems as discussed above and the idea of constructing a specification test for diffusion models comes up naturally.

To construct a test of diffusion based on infinitesimal operator, I consider a transformation based on the celebrated "martingale problems". This transformation gives us a martingale characterization for diffusion processes which is not only a complete identification but also very simple and convenient to check. Let me first define the martingale problem(see, Karatzas and Shreve(1991), Ch5.4):

Definition 4.1.5: A probability measure P on $\left(C[0,\infty)^d, \mathcal{B}(C[0,\infty)^d)\right)$ under which

$$M_t^f = f(X_t) - f(X_0) - \int_0^t (\mathcal{A}f)(X_s)ds \tag{4.13}$$

is a martingale for every $f \in D(\mathcal{A})$, is called a solution to the martingale problem associated with the operator $\mathcal{A}$.

How is the martingale problem related to the SDE-diffusion? As we know, SDE has two types of solutions: strong solutions and weak solutions(see Karatzas and Shreve(1991), Ch5.2-3 or Rogers and Williams(2000), ChV.2-3 for details). Intuitively, the strong solution is a solution to SDE with *a.s.*properties and a weak solution is the solution to SDE with in law properties. When the drift and diffusion terms of a SDE satisfy the Lipschitz and linear growth conditions, there is a strong solution to the SDE. But for general drift and diffusion terms, a strong solution may not exist;in this case, probabilists usually attempt to solve the SDE in the "weak" sense of finding a solution with the right probability law. The martingale problem is a variation of this "weak solution approach" developed by Strook and Varadhan (1969) and is in fact equivalent to the weak solution of a SDE as shown by the following:

Theorem 4.1.1: The process $\{X_t\}$ is a weak solution to the SDE (4.1) if and only if it satisfies the martingale problem of Definition 4.1.5 with $\mathcal{A}$ as the infinitesimal operator of $\{X_t\}$ defined in (4.10).

Now we have shown that the weak solution of a SDE is equivalent to the martingale problem. When strong solution exists the weak solution will coincide with it. Hence it is enough to consider the weak solution identification for doing econometric inference because regularity conditions for the existence of strong solution are usually satisfied and thus imposed in analysis[3](see Protter 2005 for some regularity Lipschitz conditions for the existence and uniqueness of a strong solution to a SDE).

By Theorem 4.1.1 and (4.13), the correct specification of a SDE-diffusion is

[3]I thank Professor Philip Protter for suggesting this point to me.

equivalent to whether the martingale problem is satisfied, implying that the hypotheses of interest H_0 in (4.3) versus H_A in (4.4) can be equivalently written as:

H_0 : For some $\theta_0 \in \Theta$, $M_t^f(\theta_0) = f(X_t) - f(X_0) - \int_0^t (\mathcal{A}_{\theta_0} f)(X_s)ds$ is a martingale for every $f \in D(\mathcal{A})$, where

$$\begin{aligned}
\mathcal{A}_{\theta_0} f(x) &= \sum_{i=1}^{d} b_i(x;\theta_0) f_i'(x) + \frac{1}{2}\sum_{i,j=1}^{d} a_{i,j}(x;\theta_0) f_{i,j}''(x) \\
a_{ij}(x;\theta_0) &= \sum_{k=1}^{d} \sigma_{i,k}(x;\theta_0)\sigma_{j,k}(x;\theta_0)
\end{aligned} \tag{4.14}$$

Versus

H_A : For all $\theta \in \Theta$, $M_t^f(\theta) = f(X_t) - f(X_0) - \int_0^t (\mathcal{A}_{\theta} f)(X_s)ds$ is not a martingale for some $f \in D(\mathcal{A})$, where

$$\begin{aligned}
\mathcal{A}_{\theta} f(x) &= \sum_{i=1}^{d} b_i(x;\theta) f_i'(x) + \frac{1}{2}\sum_{i,j=1}^{d} a_{i,j}(x;\theta) f_{i,j}''(x) \\
a_{ij}(x;\theta) &= \sum_{k=1}^{d} \sigma_{i,k}(x;\theta)\sigma_{j,k}(x;\theta)
\end{aligned} \tag{4.15}$$

Now we have transformed the correct specification hypothesis of a multivariate time-homogeneous diffusion into a martingale hypothesis for some new processes based on the infinitesimal operator and martingale problems which is very convenient to check. Observe from (4.14) that what we have to do is only to check the martingale property for the transformed processes M_t^f for every $f \in D(\mathcal{A})$. However, there are usually an infinite number of functions $f(\cdot)$ in the domain $D(\mathcal{A})$ which are usually called test functions(note that $D(\mathcal{A})$ contains the function space $C_c^\infty(int(E))$ as a subset for Feller diffusion defined in Definition4). Hence we unfortunately have to check the martingale property for infinitely many processes $\{M_t^f\}$ for test function $f \in D(\mathcal{A})$. It is definitely impossible in practice and we need a subclass of $D(\mathcal{A})$ which not only consists of finitely many function forms but also plays the same role as $D(\mathcal{A})$ does. Luckily, the following celebrated theorem gives an equivalent subclass by which a practical test procedure can be constructed easily[4].

Theorem 4.1.2: The process $\{X_t\}$ is a weak solution to the SDE in (4.1) if it satisfies the martingale problem of Definition 4.1.5 with $\mathcal{A}$ as the infinitesimal operator of $\{X_t\}$ for the choices $f(x) = x_i$ and $f(x) = x_i x_j$ with $1 \le i, j \le d$.

At first glance, this result may appear confusing because $f(x) = x_i$ and $f(x) = x_i x_j$ do not belong to $D(\mathcal{A})$ which is a subset of $C_0(\mathbb{R}^d)$. To get an intuition for this important result, let me choose sequences $\{g_i^{(K)}\}_{K=1}^{\infty}$ and $\{g_{ij}^{(K)}\}_{K=1}^{\infty}$ in function space $C_0(\mathbb{R}^d)$ such that $g_i^{(K)}(x) = x_i$ and $g_{ij}^{(K)} = x_i x_j$ for $\|x\| \le K$. If $M^{g_i^{(K)}}$ and $M^{g_{ij}^{(K)}}$ are martingales, then M^{x_i} and $M^{x_{ij}}$ are local martingales. A similar result to Theorem1 with local martingale replacing martingale then tells us that $\{X_t\}$ is a weak solution to the SDE in (4.1). Of course, the converse of Theorem 4.1.2 only holds with local martingale replacing martingale. However, since examples which are local martingales but not martingales are few and too artificial in certain sense even when they exist[5], I regard them as almost the same and do not pay much attention to their difference in this study[6]. Actually, by imposing certain regularity conditions, a local martingale can become a martingale(see

[4]We can also reduce the space of test functions to an equivalent subclass by the method considered in Kanaya(2007) which is based on the concept of a core and "approximation" theory. Since my reduced space of test functions constructed by Theorem2 is much more simple and intuitive than that in Kanaya(2007), I do not use that method here. Also see Hansen and Scheinkman(1995) and Conley, Hansen, Luttmer and Scheinkman(1997) for more discussions about choices of test functions.

[5]See Karatzas and Shreve(1991), p.168 and 200-201 for some examples which are local martingales but not martingales.

[6]When the difference really matters, the local martingale property can be used in the specification testing of diffusion models. The idea is to use the fact that the time-changed continuous local martingale by quadratic variation is a standard Brownian Motion(see Andersen, Bollerslev & Dobrev(2007) and Park(2008) for details). Since this approach is closely related to time-dependent diffusion models and the test procedure will be very different, I do not pursue it here. But the research on it is being investigated and will be reported soon.

Protter 2005 for such technical conditions). I do not explore them here since it is not the focus of this study and could certainly distract the attention. To sum up, Theorem2 implies that the hypotheses of interest H_0 in (4.14) versus H_A in (4.15) can be equivalently written as:

H_0 : For some $\theta_0 \in \Theta$

$$\begin{aligned} M_t^{x_i}(\theta_0) &= X_t^i - X_0^i - \int_0^t b_i(X_s;\theta_0)ds \\ M_t^{x_i,x_j}(\theta_0) &= X_t^{x_i,x_j} - X_0^{x_i,x_j} - \int_0^t \Big[b_i(X_s;\theta_0)X_s^j + b_j(X_s;\theta_0)X_s^i \\ &\quad + \sum_{k=1}^{d} \sigma_{i,k}(X_s;\theta_0)\sigma_{j,k}(X_s;\theta_0) \Big] ds \end{aligned} \tag{4.16}$$

are martingales for $1 \leq i, j \leq d$.

Versus

H_A : For all $\theta \in \Theta$,

$$M_t^{x_i}(\theta) or M_t^{x_i,x_j}(\theta) \tag{4.17}$$

is not a martingale for some i,j=1,$\cdots$,d, where $M_t^{x_i}(\theta)$ and $M_t^{x_i,x_j}(\theta)$ are defined as in (4.16) with θ replacing θ_0.

This greatly simplifies the hypothesis and makes the testing of the specification completely practical. Note that my hypothesis of correct specification can be expressed explicitly by the drift and diffusion terms. Therefore, any specification of the diffusion model can be tested directly without computation of transition density and the asymptotic distribution is completely free of estimation uncertainty as long as the estimator is $\sqrt{n}$-consistent. In contrast, the transition density based methods like Hong and Li (2005) or Ait-Salalia, Fan and Peng (2008) have to approximate the model-implied transition density because the transition density hardly has a closed-form.

To see the information contained in different transformed processes and prepare for the discussions of separate inference in Section 4.3 illustrated using univariate diffusion models, we state the specification hypothesis corresponding to (4.16)-(4.17) for univariate models with infinitesimal operator defined by (4.11):

H_0 : For some $\theta_0 \in \Theta$

$$
\begin{aligned}
M_t^x(\theta_0) &= X_t - X_0 - \int_0^t b(X_s;\theta_0)ds \\
M_t^{x^2}(\theta_0) &= X_t^2 - X_0^2 - \int_0^t \left[2b(X_s;\theta_0)X_s + \sigma^2(X_s;\theta_0)\right] ds
\end{aligned} \tag{4.18}
$$

are both martingales versus H_A: For all $\theta \in \Theta$,

$$M_t^x(\theta) or M_t^{x^2}(\theta) \tag{4.19}$$

is not a martingale, where $M_t^x(\theta)$ and $M_t^{x^2}(\theta)$ are defined as in (4.18) with θ replacing θ_0. For the convenience of constructing a test procedure, I further state the following equivalent hypotheses of correct specification in terms of the *m.d.s.*property for the transformed processes.

H_0 : For some $\theta_0 \in \Theta$, $E\left[Z_t(\theta_0)|\mathcal{I}_{t'}\right] = 0$ for any $t' < t$, where callI$_{t'} = \sigma\{X_{t''}\}_{t''<t'}$ is the sigma-field generated by the past information of $\{X_t\}$ at time t' and $Z_t(\theta_0)$ is a vector with components for $i, j = 1, \cdots, d$

$$
\begin{aligned}
Z_t^i(\theta_0) &= M_t^{x_i}(\theta_0) - M_{t-\Delta}^{x_i}(\theta_0) \\
&= X_t^i - X_{t-\Delta}^i - \int_{t-\Delta}^t b_i(X_s;\theta_0)ds \\
Z_t^{i,j}(\theta_0) &= M_t^{x_i x_j}(\theta_0) - M_{t-\Delta}^{x_i x_j}(\theta_0) \\
&= X_t^i X_t^j - X_{t-\Delta}^i X_{t-\Delta}^j - \int_{t-\Delta}^t \Big[b_i(X_s;\theta_0)X_s^j + b_j(X_s;\theta_0)X_s^i \\
&\quad + \sum_{k=1}^d \sigma_{i,k}(X_s;\theta_0)\sigma_{j,k}(X_s;\theta_0)\Big] ds
\end{aligned} \tag{4.20}
$$

versus

$$H_A: E\left[Z_t(\theta)|\mathcal{I}_{t'}\right] \neq 0 \tag{4.21}$$

for any $t' < t$ and all $\theta \in \Theta$, where $\mathcal{I}_{t'}$ and $Z_t(\theta)$ is defined as in (4.20) with θ replacing θ_0.

Corresponding to (4.20) and (4.21) for multivariate diffusion models, the *m.d.s.*representation of specification hypothesis for univariate case is:

H_0 : For some $\theta_0 \in \Theta$, $E\left[Z_t(\theta_0)|\mathcal{I}_{t'}\right] = 0$ for any $t' < t$, where $Z_t(\theta_0) = \left(Z_t^x(\theta_0), Z_t^{x^2}(\theta_0)\right)'$, $\mathcal{I}_{t'}$ is defined as in (4.20), and

$$\begin{aligned}
Z_t^x(\theta_0) &= M_t^x(\theta_0) - M_{t-\Delta}^x(\theta_0) \\
&= X_t - X_{t-\Delta} - \int_{t-\Delta}^t b(X_s;\theta_0)ds \\
Z_t^{x^2}(\theta_0) &= M_t^{x^2}(\theta_0) - M_{t-\Delta}^{x^2}(\theta_0) \\
&= X_t^2 - X_{t-\Delta}^2 - \int_{t-\Delta}^t \left[2b(X_s;\theta_0)X_s + \sigma^2(X_s;\theta_0)\right]ds
\end{aligned} \tag{4.22}$$

versus

$$H_A: E\left[Z_t(\theta)|\mathcal{I}_{t'}\right] \neq 0 \tag{4.23}$$

for any $t' < t$ and all $\theta \in \Theta$, where $\mathcal{I}_{t'}$ is defined as in (4.21) and $Z_t(\theta)$ is defined as in (4.22) with θ replacing θ_0

4.2 Test Procedure Based on Multivariate Generalized Spectral Derivative

In this section, I shall construct a test procedure of the correct specification hypotheses H_0 versus H_A in (4.20) and (4.21) for the multivariate diffusion process.

As an illustration, I shall also present the test procedure for H_0 versus H_A in (4.22) and (4.23) for univariate diffusion process which is a special case of that for multivariate case. The sample data is discrete in time, i.e., $\{X_{\tau\Delta}\}_{\tau=1}^{n}$ observed over a time span T with sampling interval Δ and sample size $n = T/\Delta$. Therefore, the process is in continuous time but the data sample is discrete. This is a general problem in continuous-time series econometrics not only for testing but also for estimation(see Lo (1988) and Ait-Sahalia (1996a,b) for discussions about the estimation of the discretized version of a continuous-time model). Like Ait-Sahalia (1996b), Hong and Li (2005) or Kanaya (2007), I will consider the discrete time implications of the *m.d.s.* property which is derived in continuous time[7]. The asymptotic scheme I use here is $n = T/\Delta \to \infty$. It can be obtained by either infill ($\Delta \to 0$) or long span ($T \to \infty$)[8] instead of both and this implies that my test procedure can be applied to both high-frequency and low-frequency data. In contrast, many other papers like Stanton (1997), Bandi and Philips (2003), and Kanaya (2007) assume $\Delta \to 0$ and hence can only be used for high-frequency data.

The null hypothesis is that $E\left[Z_t(\theta_0)|\mathcal{I}_{t'}\right] = 0$ for any $t' < t$, where call$\mathrm{I}_{t'}$=$\sigma\{X_{t''}\}_{t''<t'}$ is the sigma-field generated by the past information of $\{X_t\}$ and $Z_t(\theta_0)$ is a vector with components defined in (4.20). Also by (4.20), this implies

$$E\left[Z_t(\theta_0)|\mathcal{I}_{t'}^{Z}\right] = 0 \tag{4.24}$$

for any $t' < t$, where $\mathcal{I}_{t'}^{Z} = \sigma\{Z_{t''}(\theta_0)\}_{t''<t'}$ is the sigma-field generated by past information of $\{Z_t(\theta_0)\}$[9]. Since the sample data we have is $\{X_t, t = \tau\Delta\}_{\tau=0}^{n}$ with

[7]The discretization can be justified by Zahle(2008) who proves rigourously that the discrete time processes solving the disctete analogue of the martingale problem approximate weakly the solution of the stochastic differential eqnarray under additional assumption on the moments of the increments.

[8]Bandi and Phillips(2003) argued that both the infill and long-span assumptions are needed to estimate continuous-time (diffusion) process fully non-parametrically.

[9]$\mathcal{I}_{t'}$ can still be used here and this actually simplifies the test statistic in $\widehat{M}_0(p)$ (4.32) because

$n = T/\Delta$, an application of the Law of Iterated expectation as well as (3.1) implies that $E\left[Z_{\tau\Delta}(\theta_0)|\mathcal{I}^Z_{\tau-1}\right] = 0$, where

$$\mathcal{I}^Z_{\tau-1} = \sigma\{Z_{(\tau-1)\Delta}(\theta_0), Z_{(\tau-2)\Delta}(\theta_0), \cdots, Z_\Delta(\theta_0), Z_0(\theta_0)\} \tag{4.25}$$

Observe that (4.25) is a *m.d.s.* property for discrete time process $\{Z_{\tau\Delta}(\theta_0)\}_{\tau=1}^n$ and it is derived as an implication of the *m.d.s.*property in continuous time instead of a result from the discretization of the continuous time process. In this respect, it is similar to the approaches of Ait-Sahalia (1996a,b) and Lo (1988) which deal with estimation problems and therefore is free of the discretization errors which are discussed in Lo(1988) in the context of estimation. Moreover, this property is also the reason why my test procedure based on (4.25) is only assuming $n = T/\Delta \to \infty$ for asymptotic theory and applicable to both low and high frequency data. In contrast, other procedures using the discretization scheme of a continuous time model only apply to high frequency data and henceforth a little restricted(for example, see Gao and Casas(2008) and Fan and Zhang (2003)).

As discussed above, my test procedure will be based on checking whether (4.25) is true or not. However, it is not a trivial task to check this. First, the conditioning information set $\mathcal{I}^Z_{\tau-1}$ has an infinite dimension as $\tau \to \infty$ and then there is a "curse of dimensionality" difficulty associated with testing the *m.d.s.*property. Second, $\{Z_{\tau\Delta}(\theta_0)\}$ may display serial dependence in its higher order conditional moments. Any test should be robust to time-varying conditional heteroscedasticity and higher order moments of unknown form in $\{Z_{\tau\Delta}(\theta_0)\}$. To check the *m.d.s.*property of $\{Z_{\tau\Delta}(\theta_0)\}$, I extend Hong's (1999) generalized spectral approach to a multivariate generalized spectral derivative method. The idea is similar

$\{X_t\}$ is only d-dimensional while $\{Z_t\}$ is a $2d'$-dimensional process. The test procedure constructed this way can be called a generalized cross-spectral derivative approach. I do not follow this method here mainly to simplify the notations. And because of the close relationship between $\{X_t\}$ and $\{Z_t\}$ which can be seen from (2.20), it will not matter too much for the final result.

to Hong and Lee (2005) which considers testing time series conditional mean models with no prior knowledge of possible alternatives. The difference is that here the process I check for *m.d.s.*property is transformed explicitly from the original process while the process Hong and Lee (2005) check is the estimated residuals from a conditional mean model. Furthermore, the process $\{Z_{\tau\Delta}(\theta_0)\}$ is multivariate but that in Hong and Lee (2005) is only univariate. Therefore, the problem here is more complicated and we need to extend the generalized spectral approach to an multivariate one while keeping the property of being free of "curse of dimensionality". This can be regarded as another contribution of this paper.

Suppose $\{Z_\tau\}$ is a strictly stationary process with marginal characteristic function $\varphi(u) = E(e^{iu'Z_\tau})$ and pairwise joint characteristic function $\varphi_m(u,v) = E(e^{iu'Z_\tau + iv'Z_{\tau-|m|}})$, where $i = \sqrt{-1}$, $u, v \in \mathbb{R}^{d'}$, and $m = 0, \pm1, \cdots$. The basic idea of the generalized spectrum is to consider the spectrum of the transformed series $\{e^{iu'Z_\tau}\}$. It is defined as

$$f(\omega, u, v) \equiv \frac{1}{2\pi}\sum_{m=-\infty}^{\infty}\sigma_m(u,v)e^{-im\omega}, \omega \in [-\pi, \pi]$$

where ω is the frequency, and $\sigma_m(u,v)$ is the covariance function of the transformed series:

$$\sigma_m(u,v) \equiv cov(e^{iu'Z_\tau}, e^{iv'Z_{\tau-|m|}}), m = 0, \pm1, \cdots \tag{4.26}$$

Note that the function $f(\omega, u, v)$ is a complex-valued scalar function although Z_τ is a $d' \times 1$ vector. It can capture any type of pairwise serial dependence in $\{Z_\tau\}$, i.e., dependence between Z_τ and $Z_{\tau-m}$ for any nonzero lag m, including that with zero autocorrelation. First, this is analogous to the higher order spectra (Brillinger and Rosenblatt, 1967a,b) in the sense that $f(\omega, u, v)$ can capture the serial dependence in higher order moments. However, unlike the higher order spectra, $f(\omega, u, v)$ does not require existence of any moment of $\{Z_\tau\}$. This is important in economics and finance because it has been argued that the higher order moments of many financial time series may not exist. Second, this can capture nonlinear dynamics while maintaining the nice features of spectral analysis, especially its appealing property to accommodate information in all lags. In the present context, it can check the *m.d.s.*property over many lags in a pairwise manner, avoiding the "curse of dimensionality" difficulty. This is not achievable by other existing tests in the literature which only check a fixed lag order.

The generalized spectrum $f(\omega, u, v)$ itself cannot be applied directly for testing H_0, because it will capture the serial dependence not only in mean but also in higher order moments. However, just as the characteristic function can be differentiated to generate various moments of $\{Z_\tau\}$, $f(\omega, u, v)$ can be differentiated to capture the serial dependence in various moments. To capture(and only capture) the serial dependence in conditional mean, one can consider the derivative:

$$f^{(0,1,0)}(\omega, 0, v) \equiv \frac{\partial}{\partial u} f(\omega, u, v)|_{u=0} = \frac{1}{2\pi} \sum_{m=-\infty}^{\infty} \sigma_m^{(1,0)}(0, v) e^{-im\omega}, \omega \in [-\pi, \pi]$$

where

$$\sigma_m^{(1,0)}(0, v) \equiv \frac{\partial}{\partial u} \sigma_m(u, v) \mid_{u=0} = cov(iZ_\tau, e^{iv' Z_{\tau-|m|}}) \tag{4.27}$$

is a $d' \times 1$ vector. The measure $\sigma_m^{(1,0)}(0, v)$ checks whether the autoregression function $E[Z_\tau \mid Z_{\tau-m}]$ at lag order m is zero. Under some regularity conditions, $\sigma_m^{(1,0)}(0, v) = 0$ for all $v \in \mathbb{R}^{d'}$ if and only if $E[Z_\tau \mid Z_{\tau-m}] = 0$, *a.s.*

It should be noted that the hypothesis of $E\left[Z_\tau(\theta) \mid \mathcal{I}_{\tau-1}\right] = 0$ *a.s.*is not exactly the same as the hypothesis of $E[Z_\tau \mid Z_{\tau-m}] = 0$ *a.s.* for all $\tau > 0$. The former implies the latter but not vice versa. There exists a gap between them. This is the price we have to pay to deal with the difficulty of the "curse of dimensionality". Nevertheless, the examples for which $E[Z_\tau \mid Z_{\tau-m}] = 0$ *a.s.* for all $\tau > 0$ but $E\left[Z_\tau(\theta) \mid \mathcal{I}_{\tau-1}\right] \neq 0$ *a.s.*may be rare in practice and are thus pathological. Even in cases for which the gap does matter, it can be further narrowed down by using the function $E[Z_\tau \mid Z_{\tau-m}, Z_{\tau-l}]$ which may be called the bi-autoregression function of Z_τ at lags (m, l). An equivalent measure is the generalized third order central cumulant function $\sigma_{m,l}^{(1,0)}(0, v) = cov\left[Z_\tau, \exp\left(iv_1' Z_{\tau-m} + iv_2' Z_{\tau-l}\right)\right]$, where $v = (v_1, v_2) \in \mathbb{R}^{d'} \times \mathbb{R}^{d'}$.

In the present context, I suppress Δ and θ_0 and then let $Z_\tau \equiv Z_{\tau\Delta}(\theta_0)$ for the simplification of notations. Obviously, Z_τ cannot be observed. We can first estimate the parameter θ_0 by the random sample $\{X_{\tau\Delta}\}_{\tau=1}^n$ to get a $\sqrt{n}$-consistent estimator. Then the estimated processes $\widehat{Z_\tau} = Z_{\tau\Delta}(\widehat{\theta})$ is obtained. Examples of $\widehat{\theta}$ are approximated transition density based estimator in Ait-Sahalia (2002b), simulated MLE in Pedersen (1995) and so on. Then we can estimate $f^{(0,1,0)}(\omega, 0, v)$ for process $\{Z_\tau(\theta)\}$ by the following smoothed kernel estimator:

$$\widehat{f}^{(0,1,0)}(\omega, 0, v) \equiv \frac{1}{2\pi} \sum_{m=1-n}^{n-1} (1 - |m|/n)^{1/2} k(m/p) \widehat{\sigma}_m^{(1,0)}(0, v) e^{-im\omega},$$

$\omega \in [-\pi, \pi]$ and $v \in \mathbb{R}^{d'}$, where $\widehat{\sigma}_m^{(1,0)}(0, v) \equiv \frac{\partial}{\partial u} \widehat{\sigma}_m(u, v) \mid_{u=0}$, $\widehat{\sigma}_m(u, v) = \widehat{\varphi}_m(u, v) - \widehat{\varphi}_m(u, 0)\widehat{\varphi}_m(0, v)$, and

$$\widehat{\varphi}_m(u, v) = \frac{1}{n - |m|} \sum_{\tau=|m|+1}^{n} e^{iu'\widehat{Z}_\tau + iv'\widehat{Z}_{\tau-|m|}} \tag{4.28}$$

Here, $p = p(n)$ is a bandwidth, and $k : \mathbb{R} \to [-1, 1]$ is a symmetric kernel. Examples of $k(\cdot)$ include Bartlett, Daniell, Parzen and Quadratic spectral kernels(e.g., Priestley 1981, p.442). The factor $(1 - |m|/n)^{1/2}$ is a finite-sample correction and could be replaced by unity. Under certain conditions, $\widehat{f}^{(0,1,0)}(\omega, 0, v)$ is consistent for $f^{(0,1,0)}(\omega, 0, v)$.

Under H_0, we have $\sigma_m^{(1,0)}(0, v) = 0$ for all $v \in \mathbb{R}^{d'}$ and all $m \neq 0$. Consequently, the generalized spectral derivative $f^{(0,1,0)}(\omega, 0, v)$ becomes a "flat spectrum" as a function of frequency ω:

$$f_0^{(0,1,0)}(\omega, 0, v) \equiv \frac{1}{2\pi}\sigma_0^{(1,0)}(0, v) = \frac{1}{2\pi} cov\left(iZ_\tau, e^{iv'Z_\tau}\right) \tag{4.29}$$

$\omega \in [-\pi, \pi]$ and $v \in \mathbb{R}^{d'}$, which can be consistently estimated by

$$\widehat{f}_0^{(0,1,0)}(\omega, 0, v) = \frac{1}{2\pi}\widehat{\sigma}_0^{(1,0)}(0, v) \tag{4.30}$$

$\omega \in [-\pi, \pi]$ and $v \in \mathbb{R}^{d'}$.

The estimators $\widehat{f}^{(0,1,0)}(\omega, 0, v)$ and $\widehat{f}_0^{(0,1,0)}(\omega, 0, v)$ converge to the same limit under H_0 and generally converge to different limits under H_A. Thus, any significant divergence between them is evidence of the violation of the MDS property and hence of the mis-specification of the process. We can measure the distance between $\widehat{f}^{(0,1,0)}(\omega, 0, v)$ and $\widehat{f}_0^{(0,1,0)}(\omega, 0, v)$ by quadratic form:

$$\begin{aligned}
\widehat{Q} &\equiv \int\int_{-\pi}^{\pi} \left\|\widehat{f}^{(0,1,0)}(\omega, 0, v) - \widehat{f}_0^{(0,1,0)}(\omega, 0, v)\right\|^2 d\omega dW(v) \\
&= \sum_{m=1}^{n-1} k^2(m/p)(1 - m/n) \int \left\|\widehat{\sigma}_m^{(1,0)}(0, v)\right\|^2 dW(v)
\end{aligned} \tag{4.31}$$

where the second equality follows by Parseval's identity and $W(v) =_{c=1}^{d'} W_0(v_c)$ with $W_0 : \mathbb{R} \rightarrow \mathbb{R}^+$ a nondecreasing weighting function that weighs sets symmetric about the origin equally. Examples of $W_0(\cdot)$ include the CDF of any symmetric probability distribution, either discrete or continuous.

My proposed omnibus test statistic for correct specification hypothesis is an appropriately standardized version of $\widehat{Q}$,

$$\widehat{M}_0(p) = \left[\sum_{m=1}^{n-1} k^2(m/p)(n - m) \int \left\|\widehat{\sigma}_m^{(1,0)}(0, v)\right\|^2 dW(v) - \widehat{C}_0(p)\right] / \sqrt{\widehat{D}_0(p)}$$

where

$$\widehat{C}_0(p) = \sum_{m=1}^{n-1} k^2(m/p) \frac{1}{n-m} \sum_{\tau=m+1}^{n-1} \left\|\widehat{Z}_\tau\right\|^2 \int \left|\widehat{\psi}_{\tau-m}(v)\right|^2 dW(v)$$

$$\begin{aligned}
&\widehat{D}_0(p) \\
= \; & 2\sum_{m=1}^{n-2}\sum_{l=1}^{n-2} k^2(m/p)k^2(l/p) \sum_{a=1}^{d'}\sum_{a'=1}^{d'} \int\int \\
&\times \left| \frac{1}{n-\max(m,l)} \sum_{\tau=\max(m,l)+1}^{n} \left[\widehat{Z}_{a\tau}\widehat{Z}_{a'\tau}\right] \widehat{\psi}_{\tau-m}(v)\widehat{\psi}^*_{\tau-l}(u) \right|^2 dW(u)dW(v) \quad (4.32)
\end{aligned}$$

and $\widehat{\psi}_\tau(v) = e^{iv'\widehat{Z}_\tau} - n^{-1}\sum_{\tau=1}^{n} e^{iv'\widehat{Z}_\tau}$. Throughout, all unspecified integrals are taken on the support of $W(\cdot)$. The factors $\widehat{C}_0(p)$ and $\widehat{D}_0(p)$ are approximately the mean and the variance of quadratic form $n\widehat{Q}$. The impact of conditional heteroscedasticity and other time-varying higher order conditional moments has already been taken into account. Note that $\widehat{M}_0(p)$ involves d'- and $2d'$-dimensional numerical integrations which can be computationally cumbersome when d' is large. In practice, one may choose a finite number of grid points symmetric about zero or generate a finite number of points drawn from a uniform distribution on $[-1,1]^{d'}$. The asymptotic theory allows for both discrete and continuous weighting function for $W_0(\cdot)$ which weigh sets symmetric about zero equally. A continuous weighting function for $W_0(\cdot)$ will ensure good power for $\widehat{M}_0(p)$, but there is a trade-off between computational cost and power gains when choosing a discrete or continuous weighting function. One may expect that the power of $\widehat{M}_0(p)$ will be ensured if sufficiently fine grid points are used.

4.3 Separate Inference

When a model is rejected using the test procedure above which checks jointly the specification of both drift and diffusion terms, it would be interesting to explore possible sources of the rejection. Specifically, is the rejection due to the misspecified form of drift function or the diffusion function? Having this information in hand, one can try other parametric forms of drift, diffusion or both. This is particularly important when economic theory provides little guidance about the specification of the drift and diffusion, which is usually the case in practice.

However, only several papers are available in this respect and most are focused on the specification of diffusion term, like Corradi & White(1999) and Li(2007). Kristensen(2008a) and Gao & Casas(2008) develop specification tests for both the drift and diffusion terms but they need to assume the correct specification of the diffusion term a priori and hence are subject to diffusion misspecification. The tests proposed in Kristensen(2008b) and Fan & Zhang(2003) as well as those suggested although not explored in Li(2007) and Bandi & Philips(2007) do have the ability to check the drift term robust to diffusion misspecification. But the gains are achieved by the cost of nonparametrically estimating the diffusion or drift term which has already been challenged seriously by Chapman & Pearson(2000) and Pristker(1998). Moreover, to do the nonparametric estimation of drift or diffusion terms, high frequency data with sampling interval going to zero is needed(see Stanton 1997) which may not be a valid assumption for daily interest rate data. Kristensen(2008b) does not involve nonparametrically smoothing drift or diffusion term. But he relies on comparison between a semiparametric implied transition density using nonparametric smoothing for marginal density and a nonparametric directly estimated transition density. It is well known that transition density does not have a closed-form in general and hence simulation methods are used in Kristensen(2008b). Thus it is computationally burdensome and inconvenient to be applied in practice.

Since the infinitesimal operator has a closed-form in terms of drift and diffusion terms, the martingale based identification of diffusion process proposed here has the potential to do the separate inference in order to explore possible sources of the rejection. For simplicity, I only consider the univariate diffusion model with infinitesimal operator defined by (4.11) and the extension to multivariate cases is straightforward. By (2.18), the identification of the model is equivalent to the martingale property:

$$M_t^x = X_t - X_0 - \int_0^t b(X_s)ds \tag{4.33}$$

and

$$M_t^{x^2} = X_t^2 - X_0^2 - 2\int_0^t b(X_s)X_s ds - \int_0^t \sigma^2(X_s)ds \tag{4.34}$$

are both martingales.

Observe that the first transformed process M_t^x only involves the drift term and the second $M_t^{x^2}$ has both the drift and diffusion terms as inputs. Intuitively, M_t^x characterizes the dynamics of the drift term solely and this characterization is robust to the dynamics of diffusion term. Note also that $\int_0^t \sigma^2(X_s)ds$ is the so-called "integrated volatility" or the quadratic variation $[X, X]_t$ of the process $\{X_t\}$ which has received extremely intensive attention in recent years(see Andersen, Bollerslev, Diebold, and Labys, 2003; Barndorff-Nielsen and Shephard 2004, 2006; Ait-Sahalia, Mykland and Zhang 2005;). Therefore $M_t^{x^2}$ contains the dynamics of diffusion term, i.e., the volatility of the process illustrated by $\int_0^t \sigma^2(X_s)ds$. Furthermore, $M_t^{x^2}$ also characterizes the interaction between drift and diffusion terms which is represented by $\int_0^t b(X_s)X_s ds$ because $b(X_s)X_s$ will raise the power of X_s at least to 2 and hence variance will also appear in this term.

Since the characterization for the dynamics of the drift term by the martingale property of M_t^x is robust to the dynamics of diffusion term, it is conceivable that we can check the specification of the drift term robust to diffusion misspecification if we further assume the drift term is identified by this characterization (4.33). Explicitly, suppose $\{X_t\}$ follows a univariate diffusion model given by $dX_t = b(X_t, \theta)dt + \sigma(X_t)dW_t$ with $\theta \in \Theta$ and Θ a finite dimensional parameter space, then the identification assumption is

Assumption 4.3.1: There exists a unique $\theta_0 \in \Theta$ such that $M_t^x(\theta) = X_t - X_0 - \int_0^t b(X_s; \theta)ds$ is a martingale.

Actually, under this assumption (this is equivalent to Assumption 2.1 in Park (2008)), Park (2008) proposes a so-called "conditional mean model of instantaneous change for a given stochastic process"[10] which is exactly the same as M_t^x here. The difference is that his model does not consider diffusion term at all and hence it can allow a more general setup, for example, jump diffusion process and stochastic volatility models. However, Park (2008) only proposes the instantaneous conditional mean model and claims that his model covers the diffusion process as a special case while he does not provide the corresponding conditions. Suppose the underlying model is a diffusion process and we are interested in testing the specification of the drift term. Then a test based on checking the martingale property of $M_t^x(\theta)$ is not omnibus. Since the identification not only involves $M_t^x(\theta)$ in (4.33) but also involves $M_t^{x^2}$ in (4.34), it could be the case that $M_t^x(\theta)$ is a martingale but $M_t^{x^2}(\theta, \sigma(\cdot))$ is not. In such a case, the test procedure only checking the martingale property of $M_t^x(\theta)$ cannot reject the

[10]Be careful about these terminologies. The instantaneous conditional mean for continuous time stochastic processes are different from the conditional mean for discrete time models. As discussed earlier, for instance, in a general diffusion process, the conditional mean of $X_{t+\Delta}$ given X_t is usually a function not of drift solely but of both drift and diffusion terms jointly. See Ait-Sahalia (1996a) and discussions below (4.12) in this paper.

null hypothesis although it should be rejected. This under-rejection may lead to misleading conclusion about the specification of diffusion models. In other words, Assumption 4.3.1 may be too restricted as an identification assumption and may not hold in many cases if a diffusion model is considered as the underlying process.

Observe that the martingale identification of drift here is based on rigorous mathematical derivation. Therefore, my infinitesimal operator based martingale characterization actually provides the mathematical conditions that the instantaneous conditional mean in Park's (2008) model is equal to the drift a diffusion process. If there is no information about whether the underlying process is a diffusion or not, Park's (2008) model is more general while if the diffusion model is regarded as the data generating process, the infinitesimal operator based martingale characterization should be considered. Moreover Park's (2008) identification of drift can be regarded as a special case of the infinitesimal operator based martingale characterization in the case of diffusion processes. The reason is that (4.33) which is also Park's (2008) identification assumption is derived using a special choice of function forms (see Theorem 4.1.2 and discussion therein for details). If interesting function forms other than $f(x) = x_i$ and $x_i x_j$ in Theorem 4.1.2 are suitably chosen, we may get other convenient and intuitive characterizations of diffusion processes. Of course, as claimed by Park (2008), his model is a general conditional mean model of instantaneous change for continuous time stochastic processes. This makes his study more applicable in certain sense.

As a consequence, by assuming the drift term is identified by the martingale property of M_t^x, i.e., Assumption 4.3.1, a specification test for drift term can be constructed which is robust to diffusion term misspecification. The null hypothesis is the correct specification of drift term:

$$H_0 : P[b(X_t, \theta_0) = b^0(X_t)] = 1$$

for some $\theta_0 \in \Theta$ where $b^0(\cdot)$ is the true drift function, which is equivalent to

$$H_0 : M_t^x = X_t - X_0 - \int_0^t b(X_s; \theta) ds \tag{4.35}$$

is a martingale for some $\theta_0 \in \Theta$.

Following the same reasoning as that for (4.25), I can test H_0 in (4.35) by checking the following *m.d.s.*property:

$$E\left[Y_{\tau\Delta}(\theta_0) | \mathcal{I}_{\tau-1}^Y\right] = 0$$

where $\mathcal{I}_{\tau-1}^Y = \sigma\{Y_{(\tau-1)\Delta}(\theta_0), Y_{(\tau-2)\Delta}(\theta_0), \cdots, Y_\Delta(\theta_0), Y_0(\theta_0)\}$ and

$$Y_{\tau\Delta}(\theta_0) = X_{\tau\Delta} - X_{(\tau-1)\Delta} - \int_{(\tau-1)\Delta}^{\tau\Delta} b(X_s; \theta_0) ds \tag{4.36}$$

Let $Y_\tau(\theta_0) \equiv Y_{\tau\Delta}(\theta_0)$ for the simplification of notations. Obviously, $Y_\tau(\theta_0)$ cannot be observed. We first estimate the parameter θ_0 by the random sample $\{X_{\tau\Delta}\}_{\tau=1}^n$ to get a $\sqrt{n}$-consistent estimator and then the estimated processes $\widehat{Y_\tau} = Y_\tau(\widehat{\theta})$ is obtained. Since we are only interested in the specification of drift, it is better for us to use an estimation method which can estimate the parameters in the drift consistently while being robust to the diffusion misspecification. This essentially requires the estimation of a semi-parametric diffusion model with diffusion term unrestricted. Kristensen(2008a) and Ait-Sahalia(1996a) are examples of such methods. The test for checking (4.36) is a univariate special case of (4.32), i.e.,

$$\widehat{M}_1(p) = \left[\sum_{j=1}^{n-1} k^2(m/p)(n-m) \int \left|\widehat{\sigma}_m^{(1,0)}(0,v)\right|^2 dW(v) - \widehat{C}_1(p)\right] / \sqrt{\widehat{D}_1(p)}$$

where

$$\widehat{C}_1(p) = \sum_{m=1}^{n-1} k^2(m/p) \frac{1}{n-m} \sum_{\tau=m+1}^{n-1} \widehat{Y}_\tau^2 \int \left|\widehat{\psi}_{\tau-m}(v)\right|^2 dW(v)$$

$$\begin{aligned} & \widehat{D}_1(p) \\ = \; & 2 \sum_{m=1}^{n-2} \sum_{l=1}^{n-2} k^2(m/p) k^2(l/p) \\ & \int\int \left| \frac{1}{n-\max(m,l)} \sum_{\tau=\max(m,l)+1}^{n} \widehat{Y}_\tau^2 \widehat{\psi}_{\tau-m}(v) \widehat{\psi}_{\tau-l}(u) \right|^2 dW(v) dW(u) \end{aligned} \quad (4.37)$$

and $\widehat{\psi}_\tau(v) = e^{iv\widehat{Y}_\tau} - \widehat{\varphi}(v)$, and $\widehat{\varphi}(v) = n^{-1} \sum_{\tau=1}^{n} e^{iu\widehat{Y}_\tau}$ and all the terms are defined correspondingly for univariate case similar to multivariate case. Throughout, all unspecified integrals are taken on the support of $W(\cdot)$.

4.4 Asymptotic theory

4.4.1 Asymptotic distribution

Let

$$g^i(\tau,\theta) = -\int_{(\tau-1)\Delta}^{\tau\Delta} b_i(X_s;\theta) ds \quad (4.38)$$

$$\begin{aligned} g^{i,j}(\tau,\theta) = & -\int_{(\tau-1)\Delta}^{\tau\Delta} \Big[b_i(X_s;\theta) X_s^j + b_j(X_s;\theta) X_s^i \\ & + \sum_{k=1}^{d} [\sigma_{i,k}(X_s;\theta)\sigma_{j,k}(X_s;\theta)] \Big] ds \end{aligned} \quad (4.39)$$

Then we have

$$Z_\tau^i(\theta) = X_{\tau\Delta}^i - X_{(\tau-1)\Delta}^i + g^i(\tau,\theta) \quad (4.40)$$

$$Z_\tau^{i,j}(\theta) = X_{\tau\Delta}^i X_{\tau\Delta}^j - X_{(\tau-1)\Delta}^i X_{(\tau-1)\Delta}^j + g^{i,j}(\tau,\theta) \quad (4.41)$$

To derive the null asymptotic distribution the test statistic $\widehat{M}_0(p)$ in eqnarray (4.32), the following regularity conditions are imposed.

Assumption 4.4.1. $\{X_t\}$ is a strictly stationary time series such that $\mu = E[X_t]$ exists $a.s.$, and $E[||Z_\tau||^4] \leqq C$.

Assumption 4.4.2. For each sufficiently large q, there exists a strictly stationary process $\{Z_{q,\tau}\}$ measurable with respect to the sigma field generated by $\{Z_{\tau-1}, Z_{\tau-2}, \cdots, Z_{\tau-q}\}$ such that as $q \to \infty$, $Z_{q,\tau}$ is independent of $\{Z_{\tau-q-1}, Z_{\tau-q-2}, \cdots\}$ for each τ, $E[Z_{q,\tau} \mid \mathcal{I}_{\tau-1}] = 0, a.s.$where $\mathcal{I}_{\tau-1}$ is the information set at time $(\tau-1)\Delta$ that may contain lagged random variables $\{X_{(\tau-m)\Delta}, m > 0\}$ from original process and lagged random variables $\{Z_{(\tau-m)\Delta}, m > 0\}$ from the transformed process, $E\left\|Z_\tau - Z_{q,\tau}\right\|^2 \leqq Cq^{-\kappa}$ for some constant $\kappa \geqq 1$, and $E\left\|Z_{q,\tau}\right\|^4 \leqq C$ for all large q.

Assumption 4.4.3. With probability one, both $g^i(\tau, \cdot)$ and $g^{i,j}(\tau, \cdot)$ are continuously twice differentiable with respect to $\theta \in \Theta$ and $E\sup_{\theta\in\Theta}\left\|\frac{\partial}{\partial\theta}g^i(\tau,\theta)\right\|^4 \leqq C$, $E\sup_{\theta\in\Theta}\left\|\frac{\partial^2}{\partial\theta\partial\theta'}g^i(\tau,\theta)\right\|^2 \leqq C$, $E\sup_{\theta\in\Theta}\left\|\frac{\partial}{\partial\theta}g^{i,j}(\tau,\theta)\right\|^4 \leqq C$, and $E\sup_{\theta\in\Theta}\left\|\frac{\partial^2}{\partial\theta\partial\theta'}g^{i,j}(\tau,\theta)\right\|^2 \leqq C$.

Assumption 4.4.4. $\widehat{\theta} - \theta_0 = O_p(n^{-1/2})$, where $\theta_0 = p\lim(\widehat{\theta}) \in \Theta$.

Assumption 4.4.5. $k : \mathbb{R} \to [-1, 1]$ is symmetric and is continuous at $(0, 0)$ and all but a finite number of points, with $k(0) = 1$ and $|k(z)| \leqq C|z|^{-b}$ for large z and some $b > 1$.

Assumption 4.4.6. $W : \mathbb{R}^{d'} \to \mathbb{R}^+$ is nondecreasing and weighs sets symmetric about zero equally, with $\int ||v||^4 \, dW(v) \leqq C$

Assumption 4.4.7. Put $\psi_\tau(v) = e^{iv'Z_\tau} - \varphi(v)$ with $\varphi(v) = E\left[e^{iv'Z_\tau}\right]$ and $\sigma(a, a') = E\left(Z_\tau^a Z_\tau^{a'}\right)$ for $a, a' = i, ij$ and $i, j = 1, \cdots, d$(Note here ij does not denote the product between i and j but an index equivalent to (i,j). This notation applies to the whole paper). Then $\{\frac{\partial}{\partial\theta}g^i(\tau,\theta_0), Z_\tau\}$ and $\{\frac{\partial}{\partial\theta}g^{i,j}(\tau,\theta_0), Z_\tau\}$ are strictly stationary processes such that:

(a): $\sum_{m=1}^{\infty}\left\|Cov[\frac{\partial}{\partial\theta}g^a(\tau,\theta_0), \frac{\partial}{\partial\theta}g^a(\tau-m,\theta_0)]\right\| \leqq C$ for $a = i, (i,j)$ and $i, j = 1, \cdots, d$;

(b): $\sum_{m=1}^{\infty}\sup_{(u,v)\in\mathbb{R}^{2d'}}|\sigma_m(u,v)| \leqq C$;

(c): $\sum_{m=1}^{\infty}\sup_{v\in\mathbb{R}^{d'}}\left\|Cov[\frac{\partial}{\partial\theta}g^a(\tau,\theta_0), \psi_{\tau-m}(v)]\right\| \leqq C$ for $a = i, (i,j)$ and $i, j = 1, \cdots, d$;

(d): $\sum_{m,l=1}^{\infty}\sup_{(u,v)\in\mathbb{R}^{d'}}\left|E[((Z_{\tau,a}Z_{\tau,a'}) - \sigma(a,a^{'}))\psi_{\tau-m}(u)\psi_{\tau-l}(v)]\right| \leqq C$ for $a, a' = i, (i,j)$ and $i, j = 1, \cdots, d$;

(e)$\sum_{m,l,r=-\infty}^{\infty}\sup_{v\in\mathbb{R}^{d'}}\left\|\kappa_{m,l,r}(v)\right\| \leqq C$, where $\kappa_{m,l,r}(v)$ is the fourth order cumulant of the joint distribution of the process

$$\{\frac{\partial}{\partial\theta}g^a(\tau,\theta_0), \psi_{\tau-m}(v), \frac{\partial}{\partial\theta}g^a(\tau-l,\theta_0), \psi^*_{\tau-r}(v)\} \tag{4.42}$$

for $a = i, ij$ and $i, j = 1, \cdots, d$.

Assumptions 4.4.1 and 4.4.2 are regularity conditions on the data generating process (DGP). The strict stationarity on $\{X_t\}$ is imposed and the existence of the first order moment μ can be ensured by assuming $E\|X_t\|^2 < \infty$. Assumption 4.4.2 is required only under H_0. It assumes that the martingale difference sequence (*m.d.s.*) $\{Z_\tau\}$ can be approximated by a $q-$dependent *m.d.s.*process $\{Z_{q,\tau}\}$ arbitrarily well when q is sufficiently large. Because $\{Z_\tau\}$ is a *m.d.s.*, Assumption 4.4.2 essentially imposes restrictions on the serial dependence in higher order moments of X_τ. Besides, it implies ergodicity for $\{Z_\tau\}$. It holds trivially when $\{Z_\tau\}$ is a $q-$dependent process with an arbitrarily large but finite order q. In fact, this is general enough to cover many interesting processes, for example, a stochastic volatility model with short memory (see Hong and Lee (2005) for details).

Although Assumption 4.4.3 appears in terms of restrictions on $g^i(\tau, \cdot)$ and $g^{i,j}(\tau, \cdot)$, it is actually imposing moment regularity conditions on the drift and diffusion terms $b(X_\tau; \theta_0)$ and $\sigma(X_\tau; \theta_0)$ which can be seen from (4.38) and (4.39). It covers most of the popular univariate and multivariate diffusion processes in both time-homogeneous and time-inhomogeneous cases, for example, Ait-Sahalia(1996a), Ahn and Gao (1999), Chan, Karolyi, Longstaff and Sanders (1992), Cox, Ingersoll and Ross (1985), and Vasicek (1977).

Assumption A.4 requires a $\sqrt{n}$-consistent estimator $\widehat{\theta}$, which may not be asymptotically most efficient. We do not need to know the asymptotic expansion of $\widehat{\theta}$, because the sampling variation in $\widehat{\theta}$ does not affect the limit distributions of $\widehat{M}_0(p)$. This delivers a convenient and generally applicable procedure in practice, because asymptotically most efficient estimators such as *MLE* or approximated *MLE* may be difficult to obtain in practice. One could choose a suboptimal, but convenient, estimator in implementing our procedure.

Assumption 4.4.5 is a regularity condition on the kernel $k(\cdot)$. It contains all commonly used kernels in practice. The condition of $k(0) = 1$ ensures that the asymptotic bias of the smoothed kernel estimator $\widehat{f}^{(0,1,0)}(\omega, 0, v)$ in (4.28) vanishes as $n \to \infty$. The tail condition on $k(\cdot)$ requires that $k(z)$ decays to zero sufficiently fast as $|z| \to \infty$. It implies $\int_0^\infty (1 + z)k^2(z)dz < \infty$. For kernels with bounded support, such as the Bartlett and Parzen kernels, $b = \infty$. For the Daniell and quadratic-spectral kernels, $b = 1$ and 2, respectively. These two kernels have unbounded support, and thus all $(n-1)$ lags contained in the sample are used in constructing our test statistics. Assumption 4.4.6 is a condition on the weighting function $W(\cdot)$ for the transform parameter v. It is satisfied by the CDF of any symmetric continuous distribution with a finite fourth moment. Finally, Assumption 4.4.7 provides some covariance and fourth order cumulant conditions on $\{\frac{\partial}{\partial\theta}g^i(\tau,\theta_0), Z_\tau\}$ and $\{\frac{\partial}{\partial\theta}g^{i,j}(\tau,\theta_0), Z_\tau\}$, which restrict the degree of the serial dependence in $\{\frac{\partial}{\partial\theta}g^i(\tau,\theta_0), Z_\tau\}$ and $\{\frac{\partial}{\partial\theta}g^{i,j}(\tau,\theta_0), Z_\tau\}$. These conditions can be ensured by imposing more restrictive mixing and moment conditions on these two processes. However, to cover a sufficiently large class of DGPs, I choose not to do so.

I now state the asymptotic distribution of the test statistic $\widehat{M}_0(p)$ under H_0.

Theorem 4.4.1: Suppose that Assumptions 4.4.1-4.4.7 hold, and $p = cn^\lambda$ for $c \in (0,\infty)$ and $\lambda \in (0, \left(3 + \frac{1}{4b-2}\right)^{-1})$. Then under H_0,

$$\widehat{M}_0(p) \to^d N(0,1)$$

as $n \to \infty$.

As an important feature of $\widehat{M}_0(p)$, the use of the estimated processes $\{\widehat{Z_\tau}\}$ in place of the true processes $\{Z_\tau\}$ has no impact on the limit distribution of $\widehat{M}_0(p)$. One can proceed as if the true parameter value θ_0 were known and equal to $\widehat{\theta}$. The reason, as pointed out by Hong and Lee(2005), is that the convergence rate of the parametric parameter estimator $\widehat{\theta}$ to θ is faster than that of the nonparametric kernel estimator to $\widehat{f}^{(0,1,0)}(\omega,0,v)$ to $f^{(0,1,0)}(\omega,0,v)$. As a result, the limiting distribution of $\widehat{M}_0(p)$ is solely determined by $\widehat{f}^{(0,1,0)}(\omega,0,v)$ and replacing θ_0 by $\widehat{\theta}$ has no impact asymptotically. This delivers a convenient procedure, because no specific estimation method for θ_0 is required[11]. Of course, parameter estimation uncertainty in $\widehat{\theta}$ may have impact on the small sample distribution of $\widehat{M}_0(p)$. In

[11]Because of the nice properties just discussed, $\widehat{M}_0(p)$ can be used to test the *m.d.s.*hypothesis for the process with conditional heteroscedasticity of unknown form. As discussed in Hong and Lee(2005), Lobato (2002) and Park and Whang (2003) proposed some nonparametric tests of the *m.d.s.* for observed raw data using the conditioning indicator function. They also allowed for conditional heteroscedasticity, and Park and Whang (2003) allowed for nonstationary conditioning variables. However, these tests only check a fixed lag order. Moreover, their limit distributions depend on the DGP and cannot be tabulated;resampling methods have to be used

small samples, one can use a bootstrap procedure to obtain more accurate levels of the tests.

4.4.2 Asymptotic power

My tests are derived without assuming an alternative model to H_0. To gain insight into the nature of the alternatives that my tests are able to detect, I now examine the asymptotic behavior of $\widehat{M}_0(p)$ under H_A. For this purpose, a condition on the serial dependence in $\{Z_\tau\}$ is imposed:

Assumption 4.4.8. $\sum_{m=1}^{\infty} \sup_{v \in \mathbb{R}^{d'}} \left|\sigma_m^{(1,0)}(0, v)\right| \leqq C$.

Theorem 4.4.2: Suppose Assumptions 4.4.1 and 4.4.3-4.4.8 hold, and $p = cn^\lambda$ for $c \in (0, \infty)$ and $\lambda \in (0, 1/2)$. Then under H_A and as $n \to \infty$,

$$(p^{1/2}/n)\widehat{M}_0(p) \to^p \left[2D \int_0^\infty k^4(z)dz\right]^{-1/2} \sum_{m=1}^{\infty} \int \left\|\sigma_m^{(1,0)}(0, v)\right\|^2 dW(v)$$

where

$$D = 2 \sum_{a=1}^{d'} \sum_{a'=1}^{d'} E\left|Z_{a\tau} Z_{a'\tau}\right| \int \int \int_{-\pi}^{\pi} |f(\omega, u, v)|^2 \, d\omega dW(u) dW(v) \tag{4.43}$$

The constant D takes into account the impact of the serial dependence in conditioning variables $\{e^{iv'Z_{\tau-m}}\}$, which generally exists even under H_0, due to the presence of the serial dependence in the conditional variance and higher order moments of $\{Z_\tau\}$. Suppose the autoregression function $E[Z_\tau|Z_{\tau-m}] \neq 0$ at some lag $m > 0$. Then we have $\int \left\|\sigma_m^{(1,0)}(0, v)\right\|^2 dW(v) > 0$ for any weighting function $W(\cdot)$ that is positive, monotonically increasing and continuous, with unbounded

to obtain critical values on a case-by-case basis. That is why I choose to extend Hong's(1999) generalized spectral approach instead of using these methods.

support on $\mathbb{R}^{d'}$. As a consequence, $\lim_{n\to\infty} P[\widehat{M}_0(p) > C(n)] = 1$ for any constant $C(n) = o(n/p^{1/2})$ and $\widehat{M}_0(p)$ has asymptotic unit power at any given significance level, whenever $E[Z_\tau|Z_{\tau-m}] \neq 0$ at some lag $m > 0$. Note that under H_A, $\widehat{M}_0(p)$ diverges to infinity at the rate of $np^{-1/2}$, which is faster than both the rate np^d of a nonparametric transition-density based test like Ait-Sahalia, Fan and Peng (2009) and Hong and Li (2005) for $d = 1$ and the rate $np^{d/2}$ of the characteristic function based test in Chen and Hong (2010). The differences in the divergence rates can actually lead to the conclusion, by a standard proof (see Serfling (1980) for details), that the $\widehat{M}_0(p)$ test is asymptotically more powerful than the tests cited above in terms of the Bahadur (1960) asymptotic relative efficiency (ARE) under fixed alternatives[12]. Such an advantage is due to the reduction of the dimension from d to 1 by the infinitesimal operator based martingale characterization and the spectral approach for testing the *m.d.s.* Of course, it should be emphasized that the power property does not mean that the $\widehat{M}_0(p)$ test is more powerful than any other existing test against every alternative. In fact, it may be less powerful against certain specific alternatives in finite samples since a wide range of possible alternatives are incorporated. The power performances in the simulation studies of Section 4.5 show that my test is more powerful in many cases but less powerful against certain alternatives than other tests.

4.4.3 Data-Driven Lag order

A practical issue in implementing our tests is the choice of the lag order p. As an advantage, the smoothing generalized spectral approach can provide a data-

[12]The Bahadur ARE is defined as the limiting ratio of the sample sizes required by the two competing tests to attain the same asymptotic significance level under the fixed alternative models. See Serfling (1980) for details.

driven method to choose p, which, to some extent, lets data themselves speak for a proper p. Before discussing any specific method, I first justify the use of a data-driven lag order, $\widehat{p}$, say. Here, we impose a Lipschitz continuity condition on $k(\cdot)$.

Assumption 4.4.9. $|k(z_1) - k(z_2)| \leqq C\,|z_1 - z_2|$ for any (z_1, z_2) in $\mathbb{R}^2$ and some constant $C < \infty$.

This condition rules out the truncated kernel $k(z) = 1(|z| \leqq 1)$, but it still contains most commonly used nonuniform kernels.

Theorem 4.4.3: Suppose that Assumptions 4.4.1-4.4.7 and 4.4.9 hold, and $\widehat{p}$ is a data-driven bandwidth such that $\widehat{p}/p = 1 + O_p(p^{-(\frac{3}{2}\beta - 1)})$ for some $\beta > \frac{2b-1/2}{2b-1}$, where b is as in Assumption 4.4.5, and p is a nonstochastic bandwidth with $p = cn^{\lambda}$ for $c \in (0, \infty)$ and $\lambda \in (0, \left(3 + \frac{1}{4b-2}\right)^{-1})$. Then under H_0,

$$\widehat{M}_0(\widehat{p}) - \widehat{M}_0(p) \to^p 0$$

and

$$\widehat{M}_0(\widehat{p}) \to^d N(0, 1)$$

Hence, the use of $\widehat{p}$ has no impact on the limit distribution of $\widehat{M}_0(\widehat{p})$ as long as $\widehat{p}$ converges to p sufficiently fast and my test procedure enjoys an additional "nuisance parameter-free" property. Theorem6 allows for a wide range of admissible rates for $\widehat{p}$. One possible choice is the nonparametric plug-in method similar to Hong (1999, Theorem 2.2) which minimizes an asymptotic integrated mean square error (IMSE) criterion for the estimator $\widehat{f}^{(0,1,0)}(\omega, 0, v)$ in (4.28). Consider some "pilot" generalized spectral derivative estimators based on a preliminary bandwidth $\overline{p}$:

$$\overline{f}^{(0,1,0)}(\omega,0,v) = \frac{1}{2\pi}\sum_{m=1-n}^{n-1}(1-|m|/n)^{1/2}\overline{k}(m/\overline{p})\widehat{\sigma}_m^{(1,0)}(0,v)e^{-im\omega}$$

$$\overline{f}^{(q,1,0)}(\omega,0,v) = \frac{1}{2\pi}\sum_{m=1-n}^{n-1}(1-|m|/n)^{1/2}\overline{k}(m/\overline{p})\widehat{\sigma}_m^{(1,0)}(0,v)\,|m|^q\,e^{-im\omega} \tag{4.44}$$

where the kernel $\overline{k}(\cdot)$ needs not be the same as the kernel $k(\cdot)$ used in (4.28). Note that $\overline{f}^{(0,1,0)}(\omega,0,v)$ is an estimator for $f^{(0,1,0)}(\omega,0,v)$ and $\overline{f}^{(q,1,0)}(\omega,0,v)$ is an estimator for the generalized spectral derivative

$$f^{(q,1,0)}(\omega,0,v) \equiv \frac{1}{2\pi}\sum_{m=-\infty}^{\infty}\sigma_m^{(1,0)}(0,v)\,|m|^q\,e^{-im\omega} \tag{4.45}$$

For the kernel $k(\cdot)$, suppose there exists some $q \in (0,\infty)$ such that

$$0 < k^{(q)} = \lim_{z\to 0}\frac{1-k(z)}{|z|^q} \tag{4.46}$$

Then I define the plug-in bandwidth as

$$\widehat{p}_0 = \widehat{c}_0 n^{\frac{1}{2q+1}} \tag{4.47}$$

where the turning parameter estimator

$$\begin{aligned}
&\widehat{c}_0\\
=&\left\{\frac{2q(k^{(q)})^2}{\int_{-\infty}^{\infty}k^2(z)dz}\frac{\int\int_{-\pi}^{\pi}\left\|\overline{f}^{(q,1,0)}(\omega,0,v)\right\|^2 d\omega dW(v)}{\int_{-\pi}^{\pi}\left\|\int\overline{f}^{(0,1,0)}(\omega,v,-v)dW(v)\right\|^2 d\omega}\right\}^{\frac{1}{2q+1}}\\
=&\left\{\frac{2q(k^{(q)})^2}{\int_{-\infty}^{\infty}k^2(z)dz}\frac{\sum_{m=1-n}^{n-1}(n-|m|)\overline{k}^2(m/\overline{p})\,|m|^{2q}\int\left\|\widehat{\sigma}_m^{(1,0)}(0,v)\right\|^2 dW(v)}{\sum_{m=1-n}^{n-1}(n-|m|)\overline{k}^2(m/\overline{p})\widehat{R}(m)\int\left\|\widehat{\sigma}_m(v,-v)\right\|dW(v)}\right\}^{\frac{1}{2q+1}}
\end{aligned} \tag{4.48}$$

and $\widehat{R}(m) = (n - |m|)^{-1} \sum_{\tau=|m|+1}^{n} \widehat{Z}_\tau' \widehat{Z}_{\tau-|m|}$.

The data-driven $\widehat{p}_0$ in (4.47) involves the choice of a preliminary bandwidth $\overline{p}$, which can be fixed or grow with the sample size n. If it is fixed, $\widehat{p}_0$ still generally grows at rate $n^{\frac{1}{2q+1}}$ under H_A, but $\widehat{c}_0$ does not converge to the optimal tuning constant c_0 (say) that minimizes the IMSE of $\widehat{f}^{(0,1,0)}(\omega, 0, v)$ in (4.28). This is a parametric plug-in method. Alternatively, following Hong (1999), we can show that when $\overline{p}$ grows with n properly, the data-driven bandwidth $\widehat{p}_0$ in (4.47) will minimize an asymptotic IMSE of $\widehat{f}^{(0,1,0)}(\omega, 0, v)$. Simulation experiences show that the choice of $\overline{p}$ has little impact on the finite sample performances of the test; see the next section for simulation results.

4.5 Monte Carlo Simulations

In this section, I shall investigate the finite sample performances of the proposed tests $\widehat{M}_0(p)$ and $\widehat{M}_1(p)$ for joint and separate specifications respectively, with a comparison to the Hong and Li (2005) test. Since my test is constructed by a mathematical transformation and then a multivariate generalized spectral derivative approach, which pose a bit complication, I first give a clear documentation of the steps for the numerical realization to make the compaction easy to follow. Then the empirical size and power performances will be studied for both univariate and bivariate models. Last, I shall illustrate the impact of numerical approximation for the integral involved in computing the test statistics on the test performances.

4.5.1 Numerical Computation of the Tests

The computation of the tests $\widehat{M}_0(\widehat{p}_0)$ and $\widehat{M}_1(\widehat{p}_0)$ can be done by the following steps:

1. Estimate the model parameters to obtain a $\sqrt{n}$-consistent estimator $\widehat{\theta}$ for θ_0. For computing $\widehat{M}_0(\widehat{p}_0)$, a full parametric diffusion model needs to be estimated by such methods as the simulated MLE in Brandt and Santa-Clara (2002) and approximated MLE in Ait-Sahalia (2002b, 2008). But for computing $\widehat{M}_1(\widehat{p}_0)$, only drift parameters need to be estimated in a semi-parametric diffusion model and consistent estimators can be obtained by Ait-Sahalia's (1996a) OLS for univariate case with linear drift, Kristensen's (2008b) pesudo-MLE for univariate case with general drift specification and Chapter 5's conditional GMM for general multivariate models.

2. Compute the model implied processes $\{Z_\tau(\widehat{\theta})\}$ by plugging the estimator $\widehat{\theta}$ obtained in Setp 1 into (4.20) for $\widehat{M}_0(\widehat{p}_0)$ and (4.36) for $\widehat{M}_1(\widehat{p}_0)$. Note that to obtain the numerical value of the sequence $\{Z_\tau(\widehat{\theta})\}$, an integral of the $\int_{t-\Delta}^{t} f(X_s)ds$ type has to be computed. Similar to Pan (2002), I approximate these integrals by $\int_{t-\Delta}^{t} f(X_s)ds = \frac{\Delta}{2}\left[f(X_t) + f(X_{t-\Delta})\right] + O_P\left(\Delta^2\right)$. It is expected that the approximation errors should be negligible when Δ is small enough. However, it may affect the finite sample performances of the tests when the the data is sampled at very low frequency, e.g., quarterly and yearly with $\Delta = 1/4$ and 1 respectively. This is a price we need to pay by employing the infinitesimal operator which delivers many nice properties discussed above. The impact of this numerical approximation on the finite sample performances of the tests is investigated in Section 4.5.4.

3. With the estimated sequence $\{Z_\tau(\widehat{\theta})\}$, the data-driven bandwidth $\widehat{p}_0$ can be computed according to (4.47) and (4.48). Then the test statistics $\widehat{M}_0(\widehat{p}_0)$ and $\widehat{M}_1(\widehat{p}_0)$ are calculated as in (4.32) and (4.37). Since an arbitrarily preliminary bandwidth $\overline{p}$ is needed to compute $\widehat{p}_0$, I shall consider different choices of $\overline{p}$ for computing the test statistics. Simulation studies in the following show that finite sample performances of the tests do not vary much for different $\overline{p}$.

4. Finally, the test statistics $\widehat{M}_0(\widehat{p}_0)$ and $\widehat{M}_1(\widehat{p}_0)$ will be compared with the upper-tailed $N(0,1)$ critical value C_α at level α (the asymptotic critical value under the null hypothesis). If $\widehat{M}_0(\widehat{p}_0) > C_\alpha$ then reject the joint parametric specification of the model at the significant level α while for $\widehat{M}_1(\widehat{p}_0) > C_\alpha$, reject the parametric form of the drift function.

4.5.2 Empirical Size of the Test

I now study the size performances of the test procedures. To examine the size of the tests for univariate models, I simulate data from Vasicek's (1977) model (**DGP 4.5.1**):

$$dX_t = \kappa(\alpha - X_t)dt + \sigma dW_t \tag{4.49}$$

where α is the long run mean and κ is the speed of mean reversion. To illustrate the possible impact of dependent persistence in $\{X_t\}$ on the size of the test, I follow Hong and Li (2005) and Pritsker (1998) to choose two sets of parameter values, $\left(\kappa, \alpha, \sigma^2\right)$=(0.85837, 0.089102, 0.002185) and (0.214592, 0.089102, 0.000546), for the low and high persistent dependence cases respectively. The test statistic $\widehat{M}_0(p)$ is to check whether the DGP is a Vasicek model in (4.49) while the separate inference statistic $\widehat{M}_1(p)$ is for the linear drift hypothesis, i.e., $b^0(X_t) = \kappa(\alpha - X_t)$ for some κ and α.

To examine the size of the tests for multivariate models, I generate data from a Bivariate Uncorrelated Ornstein-Uhlenbeck (O-U) model (**DGP 4.5.2**), which is also the $A_0(2)$ affine diffusion term structure model in Dai and Singleton (2000):

$$d\begin{bmatrix} X_{1t} \\ X_{2t} \end{bmatrix} = \begin{bmatrix} \kappa_{11} & 0 \\ 0 & \kappa_{22} \end{bmatrix}\begin{bmatrix} X_{1t} \\ X_{2t} \end{bmatrix} dt + \begin{bmatrix} \sigma_{11} & 0 \\ 0 & \sigma_{22} \end{bmatrix} d\begin{bmatrix} W_{1t} \\ W_{2t} \end{bmatrix} \tag{4.50}$$

where W_{1t} and W_{2t} are two independent Brownian Motions and $(\kappa_{11}, \kappa_{21}, \kappa_{22}, \sigma_{11}, \sigma_{22})$ =(-0.1117, -1.1637, 1, 1). For this case, the test statistic $\widehat{M}_0(p)$ is to check whether the DGP is a Bivariate Uncorrelated O-U model in (4.50) while the separate inference statistic $\widehat{M}_1(p)$ is for the special drift specification:

$$b^0(X_t) = \begin{pmatrix} \kappa_{11}X_{1t} \\ \kappa_{22}X_{2t} \end{pmatrix} \tag{4.51}$$

for some κ_{11} and κ_{22}.

For each parameterization, we simulate 1,000 data sets of a random sample $\{X_{\tau\Delta}\}_{\tau=1}^{n}$ at the monthly frequency ($\Delta = 1/22$) for n=250, 500, and 1000 respectively. Each simulated sample path is generated using 40 intervals per month with 39 discarded out of every 40 observations, obtaining discrete observations at the monthly frequency. The simulation is carried out based on the transition density of $\{X_t\}$ which is known to be normal for both DGPs 4.5.1 and 4.5.2. These sample sizes correspond to about 20-100 years of monthly data. For each data set, we estimate the model parameters via the MLE and then compute both $\widehat{M}_0(p)$ and $\widehat{M}_1(p)$ following the steps in Section 4.5.1. The Bartlett kernel is used both in computing the data-dependent optimal bandwidth $\widehat{p}_0$ by the plug-in method for some preliminary bandwidth $\overline{p}$ and in computing the test statistics $\widehat{M}_0(\widehat{p}_0)$ and $\widehat{M}_1(\widehat{p}_0)$. The standard multivariate normal CDF is chosen for $W(\cdot)$. Simulation experiences indicate that choices of the kernel function $k(\cdot)$ and weighting function $W(\cdot)$ have no substantial impact on the size performances of tests. I consider the empirical rejection rates using the asymptotic critical values (1.28 and 1.65) at the 10% and 5% significance levels respectively.

For comparison, I describe the construction of Hong and Li (2005) test, which is based on $g(x, t|X_s, \theta)$, the model implied transition density of $X_t = x$ given X_s for $s < t$ and the true correspondent $g_0(x, t|X_s)$. For the univarite X_t, the Hong and Li (2005) test is constructed by checking the probability integral transform of the transition density

$$Q_t(\theta_0) = \int_{-\infty}^{X_t} g(x, t|X_{t-\Delta}, \theta_0)\, dx \sim i.i.d.U[0, 1]$$

under H_0. Their test is pretty robust to the persistent dependence in $\{X_t\}$ due to the transformation. However, as discussed earlier, the model implied transition density $g(x, t|X_s, \theta)$ is not in closed-form for most cases and approximation techniques are needed. More seriously, the multivariate version of the probability integral transform $Q_t(\theta_0)$ as defined above is no longer $i.i.d.U[0, 1]$ even under H_0. Although Hong and Li (2005) propose to check the multivariate diffusion models by applying the univariate $Q_t(\theta_0)$ for each state variable through a suitable partitioning, the resulting procedure does not make full use of the information for the joint dynamics of different component processes in X_t. Specifically, it may miss the misspecification in the joint dynamics of X_t for the following DGP

$$d\begin{bmatrix} X_{1t} \\ X_{2t} \end{bmatrix} = \begin{bmatrix} \kappa_{11} & 0 \\ \kappa_{21} & \kappa_{22} \end{bmatrix} \begin{bmatrix} X_{1t} \\ X_{2t} \end{bmatrix} dt + \begin{bmatrix} \sigma_{11} & 0 \\ 0 & \sigma_{22} \end{bmatrix} d\begin{bmatrix} W_{1t} \\ W_{2t} \end{bmatrix}$$

with W_{1t} and W_{2t} two independent Brownian Motions, when the model (4.50) is fit for the data. The reason is that the probability integral transforms $Q_t^1(\theta_0)$ and

$Q_t^2(\theta_0)$ for individual conditional densities $g(X_{1,t}, t|X_{t-\Delta}, X_{2,t}, \theta)$ and $g(X_{2,t}, t|X_{t-\Delta}, \theta)$ respectively, where are employed by Hong and Li (2005), are both $i.i.d.U[0,1]$ sequences while the joint dynamics are obviously misspecified due to the misspecification of the drift. Hence, the Hong and Li (2005) test has no power against such alternatives.

Table 4.1 reports the empirical sizes of $\widehat{M}_0(\widehat{p}_0)$ at the 10% and 5% levels under the correct Vasicek and Bivariate Uncorrelated O-U models. Both of the cases with low and high persistence of dependence are considered for the former. It

Table 4.1: **Empirical Sizes under DGPs 4.5.1 and 4.5.2**

	$n=250$		$n=500$		$n=1000$		$n=1500$	
	10%	5%	10%	5%	10%	5%	10%	5%
	DGP 4.5.1: High Persistent Vasicek Model							
$\widehat{M}_0(5)$	0.313	0.298	0.163	0.152	0.165	0.143	0.114	0.085
$\widehat{M}_0(10)$	0.362	0.350	0.202	0.184	0.142	0.132	0.114	0.085
$\widehat{M}_0(15)$	0.372	0.341	0.187	0.175	0.151	0.138	0.114	0.085
$\widehat{M}_0(20)$	0.325	0.274	0.168	0.146	0.162	0.145	0.114	0.085
$\widehat{M}_1(5)$	0.228	0.234	0.030	0.029	0.097	0.088	0.099	0.057
$\widehat{M}_1(10)$	0.258	0.234	0.030	0.029	0.097	0.086	0.099	0.057
$\widehat{M}_1(15)$	0.252	0.220	0.027	0.027	0.094	0.085	0.099	0.057
$\widehat{M}_1(20)$	0.243	0.213	0.030	0.028	0.097	0.088	0.102	0.057
	DGP 4.5.1: Low Persistent Vasicek Model							
$\widehat{M}_0(5)$	0.343	0.312	0.162	0.155	0.166	0.145	0.092	0.074
$\widehat{M}_0(10)$	0.380	0.373	0.210	0.189	0.145	0.133	0.092	0.074
$\widehat{M}_0(15)$	0.396	0.356	0.193	0.180	0.150	0.134	0.092	0.074

$\widehat{M}_0(20)$	0.346	0.303	0.170	0.152	0.167	0.150	0.092	0.074
$\widehat{M}_1(5)$	0.225	0.166	0.038	0.038	0.089	0.066	0.098	0.061
$\widehat{M}_1(10)$	0.225	0.166	0.038	0.035	0.082	0.064	0.098	0.061
$\widehat{M}_1(15)$	0.216	0.158	0.038	0.033	0.080	0.060	0.098	0.061
$\widehat{M}_1(20)$	0.216	0.167	0.044	0.038	0.086	0.062	0.098	0.061
	DGP 4.5.2: Bivariate Ornstein-Uhlenbeck model							
$\widehat{M}_0(5)$	0.221	0.164	0.152	0.113	0.133	0.096	0.106	0.072
$\widehat{M}_0(10)$	0.218	0.166	0.150	0.116	0.132	0.094	0.106	0.072
$\widehat{M}_0(15)$	0.206	0.165	0.152	0.120	0.132	0.094	0.106	0.070
$\widehat{M}_0(20)$	0.204	0.164	0.152	0.112	0.133	0.094	0.106	0.072
$\widehat{M}_1(5)$	0.184	0.158	0.137	0.102	0.116	0.083	0.103	0.066
$\widehat{M}_1(10)$	0.180	0.158	0.137	0.102	0.116	0.083	0.103	0.066
$\widehat{M}_1(15)$	0.182	0.157	0.135	0.100	0.114	0.080	0.101	0.066
$\widehat{M}_1(20)$	0.182	0.157	0.135	0.104	0.114	0.080	0.101	0.066

Notes : (i) 1000 iterations; (ii): DGP 4.5.1 is the Vasicek model in (4.49) with parameter values $\left(\kappa, \alpha, \sigma^2\right)$=(0.214592, 0.089102, 0.000546) and (0.85837, 0.089102, 0.002185) corresponding to high and low persistence cases respectively. DGP B0 is the Bivariate Ornstein-Uhlenbeck model in (4.50); (iii): Four choices (5, 10, 15, 20) of the preliminary bandwidth $\overline{p}$ are considered in computing $\widehat{p}_0$ with the Bartlett kernel used. The Bartlett kernel is also used for computing $\widehat{M}_0(\widehat{p}_0)$ and $\widehat{M}_1(\widehat{p}_0)$.

can be observed that there is over-rejection at both 10% and 5% levels, but the performances are improving as n increases for all three cases. Since the overrejection is still serious especially at the 5% level when n=1000, I increase the sample size to n=1500 (only for size performances) to check the empirical sizes of the tests. Obviously, when the sample size is large enough, the over-rejection is not very serious, with rejection rates around 7% at the 5% level. Furthermore, the tests display more over-rejections under strong mean reversion than under weak mean reversion. For comparison, Table 4.2 reports the empirical sizes of the Hong and Li (2005) test under the same DGPs. Similarly, the Hong and Li test has some overrejection which is close to that of the $\widehat{M}_0(\widehat{p}_0)$ tests at 10% level but much less serious at 5% level for the Vasicek model. For the Bivariate Ornstein-Uhlenbeck model, however, the Hong and Li test has more serious overrejection than my test at both 5% and 10% levels.

For the separate inference, the drift is correctly specified as a linear function for Vasicek models and as that in (4.51) for the Bivariate Uncorrelated O-U model. It can be seen that the test $\widehat{M}_1(\widehat{p}_0)$ has also nice performances for all three cases, with rejection rates around 6% at the 5% significance level when n=1000, which is actually better than the performances of $\widehat{M}_0(\widehat{p}_0)$. Therefore, the separate inference test features nice size performances. Another observation worth pointing out is that the rejection rates of both $\widehat{M}_0(\widehat{p}_0)$ and $\widehat{M}_1(\widehat{p}_0)$ do not vary much for different choices of preliminary bandwidths. This can be seen as a robust property of the optimal bandwidth based on plug-in methods.

4.5.3 Empirical Power of the Test

To investigate the power of the test for univariate diffusion models, I simulate data from the following four popular diffusion models:

Table 4.2: **Empirical Sizes and Powers of the Hong and Li (2005) test**

	$n = 250$		$n = 500$		$n = 1000$	
Models	10%	5%	10%	5%	10%	5%
	Size Performances					
DGP 4.5.1: High Persistence	0.157	0.102	0.150	0.097	0.136	0.094
DGP 4.5.1: Low Persistence	0.150	0.104	0.142	0.092	0.153	0.096
DGP 4.5.2	0.175	0.114	0.188	0.120	0.153	0.098
	Power Performances					
DGP 4.5.3	0.182	0.126	0.303	0.264	0.595	0.501
DGP 4.5.4	0.794	0.766	0.922	0.908	1.000	1.000
DGP 4.5.5	0.628	0.583	0.880	0.867	0.988	0.975
DGP 4.5.6	0.874	0.861	0.988	0.982	1.000	1.000
DGP 4.5.7	0.083	0.064	0.109	0.076	0.133	0.080
DGP 4.5.8	0.081	0.060	0.089	0.065	0.101	0.077
DGP 4.5.9	0.685	0.643	0.906	0.878	1.000	1.000

Notes : (i) 1000 iterations; (ii): DGP 4.5.1 is the Vasicek model in (4.49) with parameter values $\left(\kappa, \alpha, \sigma^2\right)$=(0.214592, 0.089102, 0.000546) and (0.85837, 0.089102, 0.002185) corresponding to high and low persistence cases respectively; DGP 4.5.2 is the Bivariate Ornstein-Uhlenbeck model in (4.50); DGPs 4.5.2-4.5.6 are CIR model, Ahn and Gao's model, CKLS model and Ait-Sahalia's nonlinear drift model, given in equations (4.52)-(4.55); DGPs 4.5.7-4.5.9 are Bivariate Correlated O-U model with constant correlations in drift and diffusion respectively and Bivariate Correlated $A_2(2)$ model in Dai and Singleton (2000)

- DGP 4.5.3 (Cox, Ingersoll, and Ross (CIR, 1985) Model):

$$dX_t = \kappa(\alpha - X_t)dt + \sigma\sqrt{X_t}dW_t \tag{4.52}$$

where $(\kappa, \alpha, \sigma^2) = (0.89218, 0.090495, 0.032742)$.

- DGP 4.5.4 (Ahn and Gao's (1999) Inverse-Feller Model):

$$dX_t = X_t[\kappa - \left(\sigma^2 - \kappa\alpha\right) X_t]dt + \sigma X_t^{3/2} dW_t \tag{4.53}$$

where $(\kappa, \alpha, \sigma^2) = (3.4387, 0.0828, 1.420864)$.

- DGP 4.5.5 (CKLS (Chan, Karolyi, Longstaff and Sanders, 1992) Model):

$$dX_t = \kappa(\alpha - X_t)dt + \sigma X_t^{\rho} dW_t \tag{4.54}$$

where $(\kappa, \alpha, \sigma^2, \rho) = (0.0972, 0.0808, 0.52186, 1.46)$.

- DGP 4.5.6 (Ait-Sahalia's (1996a) Nonlinear Drift Model):

$$dX_t = (\alpha_{-1}X_t^{-1} + \alpha_0 + \alpha_1 X_t + \alpha_2 X_t^2)dt + \sigma X_t^{\rho} dW_t \tag{4.55}$$

where $(\alpha_{-1}, \alpha_0, \alpha_1, \alpha_2, \sigma^2, \rho)$=(0.00107,-0.0517, 0.877, -4.604, 0.64754, 1.50).

Following Hong and Li (2005), the parameter values for the CIR model are taken from Pritsker (1998), and those for Ahn and Gao's model from Ahn and Gao (1999)[13]. For DGPs 4.5.5 and 4.5.6, the parameter values are taken from Ait-Sahalia's (1999) estimates of real interest rate data. For each of univariate diffusion models above, the test statistic $\widehat{M}_0(p)$ is to check whether the DGP is a Vasicek model in (4.49) while the separate inference statistic $\widehat{M}_1(p)$ is for the linear drift hypothesis, i.e., $b^0(X_t) = \kappa(\alpha - X_t)$ for some κ and α. Obviously, both of these two hypotheses should be rejected.

[13]Chen and Hong(2010) found some typos in the parameter values of Ahn and Gao's (1999) inverse-feller model by private correspondence and corrected therm. Here I choose the parameter values used by them.

To investigate the power of the test for multivariate diffusion models, sample data will be simulated from the following three bivariate models:

- DGP 4.5.7 (Bivariate Correlated O-U Model, with constant correlation in diffusion)

$$d\begin{bmatrix} X_{1t} \\ X_{2t} \end{bmatrix} = \begin{bmatrix} -0.1117 & 0 \\ 0 & -1.1637 \end{bmatrix}\begin{bmatrix} X_{1t} \\ X_{2t} \end{bmatrix} dt + \begin{bmatrix} 1 & 0 \\ 0.25 & 1 \end{bmatrix} d\begin{bmatrix} W_{1t} \\ W_{2t} \end{bmatrix} \tag{4.56}$$

- DGP 4.5.8 (Bivariate Correlated O-U Model, with constant correlation in drift):

$$d\begin{bmatrix} X_{1t} \\ X_{2t} \end{bmatrix} = \begin{bmatrix} -0.1117 & 0 \\ 0.4 & -1.1637 \end{bmatrix}\begin{bmatrix} X_{1t} \\ X_{2t} \end{bmatrix} dt + \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} d\begin{bmatrix} W_{1t} \\ W_{2t} \end{bmatrix} \tag{4.57}$$

- DGP 4.5.9 (Bivariate Correlated $A_2(2)$ model in Dai and Singleton (2000)):

$$d\begin{bmatrix} X_{1t} \\ X_{2t} \end{bmatrix} = \begin{bmatrix} -0.7 & 0.3 \\ 0.4 & -0.8 \end{bmatrix}\begin{bmatrix} X_{1t} \\ X_{2t} \end{bmatrix} dt + \begin{pmatrix} \sqrt{X_{1t}} & 0 \\ 0 & \sqrt{X_{2t}} \end{pmatrix} d\begin{bmatrix} W_{1t} \\ W_{2t} \end{bmatrix} \tag{4.58}$$

For each of bivariate diffusion models above, the test statistic $\widehat{M}_0(p)$ is to check whether the DGP is a Bivariate Uncorrelated O-U model in (4.50) while the separate inference statistic $\widehat{M}_1(p)$ is for the special drift specification in (4.51).

The perfect measure for the distances between the alternative univariate DGPs 4.5.3-4.5.6 to 4.5.1 and between the alternative bivariate DGPs 4.5.7-4.5.9 to 4.5.2 is the Kullback-Leibler information criterion since all the diffusion models have a transition density. But as discussed above, the transition density is usually not available in closed-form and hence difficult to use here for capturing the distance. Alternatively, I shall measure the distance of the model under

H_A to that under H_0 by whether the drift, diffusion or both are mis-specified. I admit that this approach may not be able to measure the exactly precise distance. However, it can give heuristic estimates for the distance between two models and moreover is very informative about separate specifications of the process dynamics.

For each of the DGPs above, I generate 1000 data sets of the random sample for $\{X_\tau\}_{\tau=\Delta}^{n\Delta}$ where n=250, 500, and 1000 at the monthly frequency, either via the transition density or Euler-Milstein scheme depending on whether the closed-form transition density is available. For DGPs 4.5.3-4.5.6, the Vasicek model implied by the null hypothesis is estimated by MLE and by OLS when the separate inference statistic is computed while for DGPs 4.5.7-4.5.9, the Bivariate Uncorrelated O-U model in (4.50) is estimated by MLE and Chapter 5's conditional GMM when computing the separate inference test statistic for each generated sample path. Then the test statistics $\widehat{M}_0(p)$ and $\widehat{M}_1(p)$ are computed following the steps in Section 4.5.1.

Table 4.3 and 4.4 report the rejection rates of $\widehat{M}_0(\widehat{p_0})$ and $\widehat{M}_1(\widehat{p_0})$ at the 10% and 5% levels for DGPs 4.5.3-4.5.6 and 4.5.7-4.5.9 respectively and for comparison, those of the Hong and Li (2005) test are reported in Table 4.2. Under DGP 4.5.3, model (4.49) is correctly specified for the drift but is misspecified for the diffusion function because it fails to capture the "level effect". The test $\widehat{M}_0(\widehat{p_0})$ has good power in this case, with rejection rates around over 96% at the 5% level when n=1000. The Hong and Li (2005) test is less powerful than the $\widehat{M}_0(\widehat{p_0})$ test, with rejection rates around 50% at the 5% level when n =1000. The separate inference test $\widehat{M}_1(\widehat{p_0})$ has also good performances with rejection rates about 9% at the 5% level when n=1000, revealing that the rejection of the model is due to

Table 4.3: **Empirical Powers Under DGPs 4.4.3-4.5.6**

	$n = 250$		$n = 500$		$n = 1000$	
	10%	5%	10%	5%	10%	5%
	DGP 4.5.3: CIR					
$\widehat{M}_0(5)$	0.656	0.634	0.778	0.764	1.000	1.000
$\widehat{M}_0(10)$	0.693	0.662	0.773	0.760	0.987	0.984
$\widehat{M}_0(15)$	0.722	0.679	0.765	0.737	0.964	0.960
$\widehat{M}_0(20)$	0.683	0.646	0.784	0.773	1.000	1.000
$\widehat{M}_1(5)$	0.483	0.426	0.139	0.123	0.152	0.120
$\widehat{M}_1(10)$	0.433	0.395	0.134	0.106	0.155	0.122
$\widehat{M}_1(15)$	0.301	0.236	0.087	0.077	0.116	0.085
$\widehat{M}_1(20)$	0.325	0.220	0.114	0.102	0.108	0.092
	DGP 4.5.4: Ahn & Gao					
$\widehat{M}_0(5)$	0.268	0.262	0.638	0.588	0.965	0.954
$\widehat{M}_0(10)$	0.252	0.247	0.635	0.583	0.967	0.962
$\widehat{M}_0(15)$	0.244	0.240	0.609	0.552	0.978	0.970
$\widehat{M}_0(20)$	0.237	0.235	0.627	0.573	0.975	0.968
$\widehat{M}_1(5)$	0.654	0.617	0.730	0.699	0.789	0.755
$\widehat{M}_1(10)$	0.632	0.580	0.734	0.692	0.763	0.730
$\widehat{M}_1(15)$	0.611	0.574	0.682	0.683	0.780	0.724
$\widehat{M}_1(20)$	0.617	0.633	0.744	0.722	0.780	0.782
	DGP 4.5.5: CKLS					
$\widehat{M}_0(5)$	0.684	0.659	0.798	0.763	1.000	1.000
$\widehat{M}_0(10)$	0.690	0.667	0.782	0.754	1.000	1.000

$\widehat{M}_0(15)$	0.716	0.701	0.764	0.752	1.000	1.000
$\widehat{M}_0(20)$	0.689	0.673	0.758	0.747	1.000	1.000
$\widehat{M}_1(5)$	0.056	0.022	0.068	0.030	0.120	0.072
$\widehat{M}_1(10)$	0.050	0.022	0.068	0.030	0.109	0.089
$\widehat{M}_1(15)$	0.043	0.018	0.060	0.030	0.109	0.060
$\widehat{M}_1(20)$	0.042	0.018	0.047	0.029	0.110	0.058
	DGP 4.5.6: Ait-Sahalia					
$\widehat{M}_0(5)$	0.475	0.425	0.838	0.802	0.967	0.962
$\widehat{M}_0(10)$	0.454	0.414	0.835	0.798	0.978	0.973
$\widehat{M}_0(15)$	0.435	0.422	0.852	0.814	0.962	0.958
$\widehat{M}_0(20)$	0.449	0.425	0.867	0.832	0.957	0.942
$\widehat{M}_1(5)$	0.432	0.403	0.798	0.754	1.000	1.000
$\widehat{M}_1(10)$	0.443	0.405	0.766	0.732	1.000	1.000
$\widehat{M}_1(15)$	0.421	0.388	0.750	0.690	1.000	1.000
$\widehat{M}_1(20)$	0.364	0.366	0.802	0.791	1.000	1.000

Notes : (i) 1000 iterations; (ii) DGPs 4.5.3-4.5.6 are CIR model, Ahn and Gao's(1999) inverse-feller model, CKLS model and Ait-Sahalia's(1996a) nonlinear drift model, given in equations (4.52)-(4.55) (iii): Four choices (5, 10, 15, 20) of the preliminary bandwidth $\overline{p}$ are considered in computing $\widehat{p}_0$ with the Bartlett kernel used. The Bartlett kernel is also used for computing $\widehat{M}_0(\widehat{p}_0)$ and $\widehat{M}_1(\widehat{p}_0)$.

the mis-specification of the diffusion term instead of the drift term.

Table 4.4: **Empirical Powers Under DGPs 4.4.7-4.5.9**

	$n = 250$		$n = 500$		$n = 1000$	
	10%	5%	10%	5%	10%	5%
	DGP 4.5.7					
$\widehat{M}_0(5)$	0.366	0.250	0.571	0.338	0.832	0.690
$\widehat{M}_0(10)$	0.368	0.252	0.570	0.340	0.830	0.696
$\widehat{M}_0(15)$	0.373	0.260	0.575	0.334	0.830	0.702
$\widehat{M}_0(20)$	0.368	0.259	0.580	0.334	0.837	0.696
$\widehat{M}_1(5)$	0.096	0.052	0.104	0.060	0.110	0.072
$\widehat{M}_1(10)$	0.094	0.052	0.104	0.060	0.109	0.075
$\widehat{M}_1(15)$	0.094	0.057	0.104	0.063	0.109	0.070
$\widehat{M}_1(20)$	0.094	0.057	0.105	0.060	0.110	0.075
	DGP 4.5.8					
$\widehat{M}_0(5)$	0.483	0.372	0.880	0.643	1.000	0.994
$\widehat{M}_0(10)$	0.480	0.372	0.884	0.647	1.000	0.995
$\widehat{M}_0(15)$	0.480	0.374	0.884	0.648	1.000	0.993
$\widehat{M}_0(20)$	0.482	0.374	0.882	0.640	1.000	0.993
$\widehat{M}_1(5)$	0.504	0.425	0.901	0.883	1.000	1.000
$\widehat{M}_1(10)$	0.506	0.426	0.900	0.890	1.000	1.000
$\widehat{M}_1(15)$	0.504	0.426	0.906	0.894	1.000	1.000
$\widehat{M}_1(20)$	0.504	0.426	0.906	0.897	1.000	1.000
	DGP 4.5.9					
$\widehat{M}_0(5)$	0.873	0.808	0.994	0.981	1.000	1.000
$\widehat{M}_0(10)$	0.870	0.810	0.994	0.980	1.000	1.000
$\widehat{M}_0(15)$	0.866	0.805	0.990	0.977	1.000	1.000

$\widehat{M}_0(20)$	0.870	0.802	0.988	0.980	1.000	1.000
$\widehat{M}_1(5)$	0.922	0.908	1.000	1.000	1.000	1.000
$\widehat{M}_1(10)$	0.920	0.908	1.000	1.000	1.000	1.000
$\widehat{M}_1(15)$	0.915	0.904	1.000	1.000	1.000	1.000
$\widehat{M}_1(20)$	0.912	0.901	1.000	1.000	1.000	1.000

Notes : (i) 1000 iterations; (ii) DGPs 4.5.7-4.5.9 are Bivariate Correlated O-U model with constant correlations in drift and diffusion respectively and Bivariate Correlated $A_2(2)$ model in Dai and Singleton (2000); (iii): Four choices (5, 10, 15, 20) of the preliminary bandwidth $\overline{p}$ are considered in computing $\widehat{p_0}$ with the Bartlett kernel used. The Bartlett kernel is also used for computing $\widehat{M}_0(\widehat{p_0})$ and $\widehat{M}_1(\widehat{p_0})$.

Under DGP 4.5.4, model (4.49) is misspecified for both the instantaneous conditional mean and variance because it ignores the nonlinear drift and diffusion. As expected, both $\widehat{M}_0(\widehat{p_0})$ and $\widehat{M}_1(\widehat{p_0})$ have excellent power when the Vasicek model (4.49) is used to fit the DGP 4.5.4. The power of $\widehat{M}_0(\widehat{p_0})$ increases substantially with the sample size n and approaches unity when n=1000 while the rejection rates of $\widehat{M}_1(\widehat{p_0})$ are around 75% at 5% level, implying the misspecification of the drift function. The Hong and Li (2005) test is more powerful than the $\widehat{M}_0(\widehat{p_0})$ tests for small sample sizes but the difference becomes negligible when n is increased to 1000.

Similar to DGP 4.4.3, DGP 4.5.5 is only mis-specified for the diffusion term, with the only difference that the coefficient of elasticity for volatility is equal to 1.46 rather than 0.5. The rejection rates of $\widehat{M}_0(\widehat{p_0})$ increases very quickly from around 65% when n=250 to over 100% when n=1000 at the 5% level. The Hong and Li (2005) test is very comparable to the $\widehat{M}_0(\widehat{p}_0)$ tests in the power performances and slightly less powerful when n=1000. For the separate inference, the rejection rates of $\widehat{M}_1(\widehat{p}_0)$ is around 7% at 5% level when n=1000, indicating the true source of rejection is the mis-specification of the diffusion function. Under DGP 4.5.6, model (4.49) is misspecified for both the drift and diffusion terms because it ignores the nonlinearity in both terms. The rejection rates are already over 80% when n=500 for $\widehat{M}_0(\widehat{p}_0)$ and 100% for $\widehat{M}_1(\widehat{p}_0)$ when n=1000 both at the 5% level. The Hong and Li (2005) test is more powerful than $\widehat{M}_0(\widehat{p}_0)$ when n=250 but the difference becomes smaller as n increases.

The results above for univariate diffusion models show that the combination of the proposed tests $\widehat{M}_0(\widehat{p}_0)$ and $\widehat{M}_1(\widehat{p}_0)$ not only have good power in detecting various model mis-specifications but also are excellent in uncovering the sources of mis-specification. In the following, I shall check their power performances for the multivariate diffusion models under DGPs 4.5.7-4.5.9. Under DGP 4.5.7, model (4.50) is correctly specified for the drift but mis-specified for the diffusion function since it misses the nonzero constant correlation in the state variables. The test $\widehat{M}_0(\widehat{p}_0)$ has good power in detecting the misspecification in the joint dynamics, with rejection rate around 70% when n=1000. In contrast, the Hong and Li (2005) test has no power and the rejection rate is only 13% at the 10% level when n=1000. This is not surprising since the conditional densities of individual sate variables are correctly specified although the joint dynamics are not. The separate inference test $\widehat{M}_1(\widehat{p}_0)$ has also good performances with rejection rate about 7% at the 5% level when n=1000, indicating that the diffusion rather than the drift function is misspecified.

Under DGP 4.5.8, model (4.50) is correctly specified for the diffusion but mis-specified for the drift function. The rejection rate of $\widehat{M}_0(\widehat{p}_0)$ increases very quickly and approaches unity when n is rising to 1000 while that of the Hong and Li (2005) test is only about 7% at the 5% level when n=1000. This confirms again that my test is powerful against misspecifications in the joint dynamics which the Hong and Li (2005) test would miss. Moreover, the separate inference test $\widehat{M}_1(\widehat{p}_0)$ has also good power against the drift misspecification, with 100% rejection rate at the 5% level when n=1000. Under DGP 4.5.9, model (4.50) is mis-specified for both the drift and diffusion functions. Both $\widehat{M}_0(\widehat{p}_0)$ and the Hong and Li (2005) test have nice power performances with the former more powerful when the sample size is only 500. The separate inference test $\widehat{M}_1(\widehat{p}_0)$ has also excellent power against the drift misspecification, with 100% rejection rate at the 5% level when n=500.

In summary, the following observations are made: (1), For both univariate and bivariate models, the $\widehat{M}_0(\widehat{p}_0)$ and $\widehat{M}_1(\widehat{p}_0)$ tests have reasonable sizes in fintie samples. (2), The $\widehat{M}_0(\widehat{p}_0)$ test has nice power against various model misspecifications. Particulary, it has excellent power in identifying misspecifications of the joint dynamics for multivariate diffusion models even when the individual component processes are correctly specified. This feature cannot be attained by the Hong and Li (2005) test which though peroforms well for univariate cases. (3), The separate inference test $\widehat{M}_1(\widehat{p}_0)$ has nice performances in revealing the sources of rejection, i.e., whether the rejection is due to drift or diffusion misspecification.

4.5.4 The Impact of Numerical Integral Approximation

As discussed earlier, the compaction of my tests involves a numerical approximation for some integrals which come from the infinitesimal operator based martingale characterization. This may affect the finite sample performances of the tests when the sampling interval Δ is not small enough, which is a price we need to pay by enjoying many nice properties such as being convenient for multivariate cases and able to check the separate specifications. To investigate the impact of the numerical integral approximation and under which frequency of the data sampling my tests are robust to the approximation errors, I shall check the size performances of the test[14] $\widehat{M}_0(\widehat{p}_0)$ by changing the sampling interval Δ.

I consider the univariate Vasicek model in (4.49) with high persistence and Bivariate Uncorrelated O-U model in (4.50). The data generating schemes are exactly the same as those in Section 4.5.2 with the exception that only n=1500 is considered and the sampling interval is set at daily ($\Delta = 1/152$), monthly ($\Delta = 1/22$), quarterly ($\Delta = 1/4$), and yearly ($\Delta = 1$) respectively. Table 4.5 reports the rejection rates of $\widehat{M}_0(\widehat{p}_0)$ at the 10% and 5% levels. It can be observed that for the Vasicek model, the test $\widehat{M}_0(\widehat{p}_0)$ has excellent size performances when the sampling frequencies are daily and monthly, with rejection rates around 6% and 8% at the 5% significance level. When the data is sampled quarterly, the test exhibits a bit overrejection but not very excessive with rejection rate about 10% at the 5% level. However, when the sampling frequency is increased to yearly, the test has serious overrejection with the rejection rate around 22% at the 5% level. Similarly for the Bivariate Uncorrelated O-U model, $\widehat{M}_0(\widehat{p}_0)$ has excellent size performances when the sampling frequencies are daily and monthly, a bit but

[14]The performances of the separate inference test $\widehat{M}_1(\widehat{p}_0)$ are also checked following the same simulation design. The performance patters are very similar to those for the $\widehat{M}_0(\widehat{p}_0)$ test.

Table 4.5: **The Impact of Numerical Integral Approximation**

	Daily(Δ=1/252)		Monthly(Δ=1/22)		Quarterly(Δ=1/4)		Yearly(Δ=1)	
	10%	5%	10%	5%	10%	5%	10%	5%
	DGP 4.5.1: Vasicek Model with High Persistence							
$\widehat{M}_0(5)$	0.103	0.058	0.114	0.085	0.152	0.107	0.260	0.228
$\widehat{M}_0(10)$	0.103	0.058	0.114	0.085	0.152	0.107	0.264	0.230
$\widehat{M}_0(15)$	0.106	0.060	0.114	0.085	0.152	0.110	0.268	0.222
$\widehat{M}_0(20)$	0.106	0.060	0.114	0.085	0.150	0.110	0.268	0.225
	DGP 4.5.2: Bivariate Ornstein-Uhlenbeck model							
$\widehat{M}_0(5)$	0.105	0.061	0.132	0.092	0.175	0.123	0.304	0.256
$\widehat{M}_0(10)$	0.105	0.057	0.132	0.090	0.175	0.120	0.298	0.254
$\widehat{M}_0(15)$	0.107	0.054	0.132	0.086	0.170	0.126	0.298	0.250
$\widehat{M}_0(20)$	0.105	0.053	0.132	0.086	0.169	0.126	0.300	0.250

Notes : (i) The iteration number is 1000 while the sample size is 1500. (ii): DGP 4.5.1 is the Vasicek model in (4.49) with high persistence and DGP 4.5.2 is the Bivariate Ornstein-Uhlenbeck model in (4.50). (iii): Four choices (5, 10, 15, 20) of the preliminary bandwidth $\overline{p}$ are considered in computing $\widehat{p}_0$ with the Bartlett kernel used. The Bartlett kernel is also used for computing $\widehat{M}_0(\widehat{p}_0)$ and $\widehat{M}_1(\widehat{p}_0)$.

not very serious overrejetion at quarterly frequency and serious overrejection when the data is sampled yearly. . When the data is sampled quarterly, the test exhibits a bit overrejection but not very excessive with rejection rate about 10% at the 5% level.

To sum up, the approximation errors for the numerical integral involved in the test statistic have serious impact on the test performances only when the sampling frequency is yearly. This is the price paid by employing the infinitesimal operator based martingale characterization which delivers many nice properties for my test procedures. As far as the data are as frequent as or higher than monthly, the approximation has little impact and the tests have nice finite sample performances. Therefore, it seems that this is not very empirically relevant since in the fields where diffusion models are used, monthly data and even data sampled higher than monthly are usually available. For example, daily or even intra-daily data can be obtained for stocks, options, and bonds in finance research. Even in the case only very low frequent data are available, this problem can be circumvented by generating higher frequent data in the sampling interval Δ similar to Brandt and Santa-Clara (2002) according to the estimated models and then compute the integrals by taking the average of the generated sample paths.

4.6 Empirical Application: Short-Rate Dynamics

In this section, I shall apply the proposed test procedure to investigate the dynamics of short-term interest rates as an empirical application[15]. The data set is the same as that in Ait-Sahalia (1996a), i.e., daily Eurodollar rates from June 1, 1973 to February 25, 1995, with a total of 5505 observations. See Ait-Sahalia (1996a) for detailed summary statistics for the data.

Five popular models are considered: the Vasicek, CIR, Ahn and Gao, CKLS, and Ait-Sahalia's nonlinear drift models, as given in (4.49)–(4.53). For each model, I estimate parameters via MLE for a full parametric model and OLS

[15] I am grateful to an anonymous referee for suggesting this empirical study.

Table 4.6: **Testing Spot Rate Models**

	Vasicek	CIR	CKLS	Ahn & Gao	Ait-Sahalia
$\widehat{M}_0(5)$	731.7	503.2	481.0	238.6	173.4
$\widehat{M}_0(10)$	728.2	509.3	476.8	230.3	180.0
$\widehat{M}_0(15)$	720.9	510.2	472.4	215.2	169.7
$\widehat{M}_0(20)$	725.5	502.4	480.3	207.2	165.5
$\widehat{M}_1(5)$	422.4	-	-	158.3	140.0
$\widehat{M}_1(10)$	430.3	-	-	150.6	142.2
$\widehat{M}_1(15)$	433.6	-	-	147.9	133.6
$\widehat{M}_1(20)$	424.7	-	-	144.8	130.3

Notes : (i): The model parameters are estimated by MLE for a full parametric model and by OLS and Chapter 5's estimator when only drift parameters are estimated. (ii): The sample period for the daily Eurodollar interest rates is from 6/01/1973 to 2/25/1995. (iii): Four choices (5, 10, 15, 20) of the preliminary bandwidth $\overline{p}$ are considered in computing $\widehat{p}_0$ with the Bartlett kernel used. The Bartlett kernel is also used for computing $\widehat{M}_0(\widehat{p}_0)$ and $\widehat{M}_1(\widehat{p}_0)$.

and Chapter 5's conditional GMM when only drift parameters are estimated. For the Vasicek, CIR, and Ahn and Gao's models, the model likelihood function has a closed-form. For the CKLS and Ait-Sahalia's nonlinear drift models, Ait-Sahalia's (2002a) closed form approximations for the model likelihood are used. With the parameter estimates in hand, the test statistic is computed following the computation procedure in Section 4.5.1.

The empirical results are reported in Table 4.6. It shows that the $\widehat{M}_0(\widehat{p}_0)$ statistics with the four choices (5, 10, 15, and 20) of preliminary bandwidths for the five models range from 165.5 to 731.7. Compared to upper-tailed N(0,1) critical values (e.g., 2.33 at the 1% level), these huge values of $\widehat{M}_0(\widehat{p}_0)$ statistics implies strong evidence that all five models are severely misspecified. Similar to Hong and Li (2005), the Vasicek model performs the worst, with the test values around 720 for all preliminary bandwidths, probably due to its restrictive assumption of constant volatility. The CIR and CKLS models dramatically reduces the test statistics values to about 500, obviously because of the more flexible diffusion specifications. The goodness of fit is further improved substantially by Ahn and Gao and Ait-Sahalia's nonlinear drift models. The latter performs the best, with the test statistic values around 170, which is the most flexible model among the five for both drift and diffusion specifications. These findings demonstrate the power of my test: they overwhelmingly reject all parametric forms, including the CKLS and Ait-Sahalia models, which Ait-Sahalia's (1996a) marginal density based test fails to reject.

To explore the sources of rejection for the spot rate models above, I report, in Table 4.6, the separate inference statistics $\widehat{M}_0(\widehat{p}_0)$ defined in (4.37) for each model. For the Vasicek, CIR and CKLS models, the statistic $\widehat{M}_0(\widehat{p}_0)$ is to check whether the drift is linear, i.e., $H_{0,1}$: $b(X_t) = \kappa(\alpha - X_t)$ for some κ and α while for the Ahn and Gao's and Ait-Sahalia's models, it is testing whether the drift follows two specific nonlinear forms, respectively: $H_{0,2}$: $b(X_t) = X_t\kappa - \alpha X_t^2$ for some κ and α and $H_{0,3}$: $b(X_t) = \alpha_{-1}X_t^{-1} + \alpha_0 + \alpha_1 X_t + \alpha_2 X_t^2$ for some $(\alpha_{-1}, \alpha_0, \alpha_1, \alpha_2)$. These tests are are actually related to the literature of the debate about whether the drift of the interest rate process is linear or not. The early studies Ait-Sahalia (1996a) and Stanton (1997) use smoothed nonparametric kernel methods to estimate the drift of the short rate and find nonlinearity, Chapman and Pearson (2000), in a striking simulation study, find that the evidence of nonlinearity documented may be spurious due to the nature of smoothed nonparametric kernel estimation. Since then, many research have appeared exploring this issue, but most of them only estimate the drift term either parametrically or non-parametrically and check if the estimated drift is linear (see, e.g., Sam and Jiang (2007) and Takamizawa (2008)), which cannot lead to a rigorous econometric procedure. In contrast, the test proposed in this study is able to check the whole dynamics of a spot rate model, reveal sources of rejection and point to the direction of a better model.

It can be seen from Table 4.6 that the linear drift hypothesis is rejected strongly with the test statistic around 420. The quadratic drift specification in $H_{0,2}$ and general nonlinear drift reduce the test statistic value dramatically to around 150 but are still rejected. These findings tell us that the drift mis-specifications do play an important role in the rejection of all five spot rate models and a potential direction for more accurate models is to consider models with nonlinear drift specifications. The latter conclusion is in sharp contrast with Hong and Li (2005) who claim that the nonlinear drift model underperforms the linear drift models based on their separate inference statistics. However, as discussed in Section 4.3, their test statistics for separate inference are only for the conditional mean for a fixed sampling interval Δ instead of for the instantaneous conditional mean or the drift with $\Delta \to 0$. Therefore, their conclusion is only valid when a discrete time model is employed to fit the short-term interest rate. In contrast, my separate inference test statistic $\widehat{M}_0(\widehat{p}_0)$ is able to check the dynamics of the drift as the instantaneous conditional mean with $\Delta \to 0$. The results in Table 4.6 show that, different from Hong and Li (2005), nonlinear drift outperforms linear drift substantially and should be an important consideration in building more accurate models for the spot rate.

4.7 Conclusion

I develop an omnibus specification test for diffusion models based on the infinitesimal operator instead of the already extensively used transition density. The infinitesimal operator-based identification of the diffusion process is equivalent to a ″martingale hypothesis″ for the new processes transformed from the original diffusion process by the celebrated ″martingale problems″. My test procedure is to check the ″martingale hypothesis″ by a multivariate generalized spectral derivative approach which has many good properties. The infinitesimal operator of the diffusion process enjoys the nice property of being a closed-form expression of drift and diffusion terms. This makes my test procedure capable of checking both univariate and multivariate diffusion models and particularly powerful and convenient for the multivariate case while in contrast checking the multivariate diffusion models is very difficult by transition density-based methods because transition density does not have a closed-form in general.

Moreover, different transformed martingale processes via the infinitesimal operator based martingale characterization contain different separate information about the drift and diffusion terms or their interactions. This motivates us to discuss several feasible test procedures which are to do separate inference to explore the sources when rejection of a parametric form happens. Finally, simulation studies show that the proposed tests have reasonable size performances and excellent power performances in finite sample.

A drawback of the infinitesimal operator based identification is that it only holds for the pure diffusion process and will fail when the sample path of the process exhibits discontinuities, the so-called "jumps". That is, my test procedure actually rules out jumps a priori. This is somewhat unsatisfactory in practice especially for high frequency data for which jumps are now believed to be an essential component of asset price dynamics both empirically and theoretically (Ait-Sahalia 2002a; Barndorff-Nielsen and Shephard 2004, 2006; Lee and Mykland 2008; Ait-Sahalia and Jacod 2008; Andersen et al. 2002; Johannes 2004; and Pan 2002). In this case, we can first identify and then discard the jump points from the sample path using methods in Lee and Mykland (2008), Andersen, Bollerslev and Dobrev (2007), Fan and Fan (2008), and Fan and Wang (2007). Such a two-step procedure extends the proposed test and is more applicable.

CHAPTER 5

ESTIMATING SEMI-PARAMETRIC DIFFUSION MODELS WITH UNRESTRICTED VOLATILITY VIA INFINITESIMAL OPERATOR BASED CHARACTERIZATION

5.1 Infinitesimal Operator Based Conditional Moment Restrictions

The model we consider in this paper is a semi-parametric time-homogeneous multivariate diffusion model, defined by the following stochastic differential eqnarray(SDE) on the filtered probability space $(\Omega, \mathfrak{F}, \{\mathfrak{F}_t\}_{t\geq 0}, P)$:

$$dX_t = b(X_t;\theta)dt + \sigma(X_t)dW_t \tag{5.1}$$

where W_t is a $d \times 1$ standard Brownian motion in $\mathbb{R}^d$, $b : E \subset \mathbb{R}^d \to \mathbb{R}^d$ is a drift function(i.e., instantaneous conditional mean), $\sigma : E \to \mathbb{R}^{d\times d}$ is a diffusion function(i.e., the instantaneous conditional standard deviation), and $\Theta \subset \mathbb{R}^q$ is a finite-dimensional parameter space. In addition, E is often called state space and we let $\mathcal{B}(E)$ be the Borel field such that $(E, \mathcal{B}(E))$ is a measurable space.

Under usual regularity conditions, $\{X_t\}$ is a continuous time Markov process with transition function $P(t, x, \Gamma) \equiv P(X_t \in \Gamma | X_0 = x)$ which is the probability that X_t, starting from the point x, is in the set Γ at time t. The Markov property is characterized by the so-called Chapman-Kolmogrov eqnarray: for $s, t \geqq 0$, $x \in E$ and $\Gamma \in \mathcal{B}(E)$, $P_{t+s}(x, \Gamma) = \int_E P_s(x, dy)P_t(y, \Gamma)$. An alternative and equivalent characterization is the induced family $\{P_t\}$ which is a set of positive bounded operators with norm less than or equal to 1 on $b(\mathcal{B}(E))$(bounded and $\mathcal{B}(E)$-measurable functions) and which is defined by:

$$P_t f(x) \equiv (P_t f)(x) = \int_E P_t(x, dy) f(y) \tag{5.2}$$

In this case, the Markov property is expressed as the following semi-group property, i.e., $P_s P_t = P_{s+t}$,for any $s, t \geqq 0$ which is equivalent to the Chapman-Kolmogrov eqnarray above. Both transition function and the semi-group of operators characterize the Markov process and interact with the sample-path property of the process. This interaction can actually be used to define the so-called Feller process which includes the diffusion process in (2.1) as a special case. Let $C_0 = C_0(E)$ be defined as the space of real-valued, continuous functions on E which vanish at infinity, i.e., $\lim_{|x|\to\infty} f(x) = 0$, equipped with the sup-norm $\|f\| \equiv \sup_{x\in E} f(x)$. By Rogers and Williams(2000, Ch III.6), a process $\{X_t\}$ is a Feller process if its semi-group of operators $\{P_t\}_{t\geqq 0}$ satisfies the following two properties: (i) $P_t C_0 \subset C_0$ for all $t \geqq 0$; (ii) for any $f \in C_0$ and $x \in E$, $P_t f(x) \to f(x)$ as $t \downarrow 0$. Feller process has good path properties[1] and is also general enough to contain most processes we are interested in, for example, diffusion processes which have been extensively used in finance and Levy processes including Poisson process and Compound process which have been receiving more and more attention in finance recently(see Schoutens(2003)).

For Feller processes, another characterization except the transition function and semi-group of operators introduced above is used more frequently in probability theory, i.e., the infinitesimal operator. It is defined as follows: A function $f \in C_0$ is said to belong to the domain $D(\mathcal{A})$ of the infinitesimal operator $\mathcal{A}$ of a Feller process X if the following limit exists:

[1]By Rogers and Williams(2000, Ch III.7-9), the canonical Feller process always admits a Cadlag(the path of the process is right continuous and has left limits) modification and satisfies the strong Markov property

$$\mathcal{A}f = \lim_{t \downarrow 0} \frac{P_t f - f}{t} \tag{5.3}$$

where $D(\mathcal{A})$ denotes the domain of $\mathcal{A}$, i.e., the family of functions(usually named test functions) in C_0 for which the limit in (5.3) exists with respect to the sup-norm of C_0[2]. Obviously, $\mathcal{A}$ is a linear operator from $D(\mathcal{A})$ to C_0. It can be seen from (5.3) immediately, that it holds P-$a.s.$for $f \in D(\mathcal{A})$

$$E\left(\frac{f(X_{t+h}) - f(X_t)}{h} | \mathfrak{F}_t\right) = \mathcal{A}f(X_t) + o(h), \tag{5.4}$$

as $h \downarrow 0$. In this sense, the infinitesimal operator indeed describes the movement of the process in an infinitesimally small time interval. Therefore, the infinitesimal operator characterizes the whole dynamics of a Feller process because the time is continuous here. In fact, it can be proved that the infinitesimal operator is equivalent to the semi-group of operators in characterizing a Feller process(see the Hill-Yoshida theorem in Dynkin(1965)). By the equivalence of semi-group of operators and the transition function, the infinitesimal operator is equivalent to the transition function in fully characterizing the dynamics of the process.

For the diffusion process in (2.1), the infinitesimal operator always has an explicit closed-form expression which can be identified by the drift and diffusion terms. According to Kallenberg(2002, Thm 19.24) and Rogers and Williams (2000, Vol1, Thm III.13.3 and Vol2, Ch V.2), the infinitesimal operator of the diffusion model in (5.1) is:

[2]Without using the sup-norm, Hansen and Scheinkman(1995) define infinitesimal operaor in the Hilber space $L^2(Q)$ where Q is an invariant(stationary) distribution of the process. This Hilber space based definition is needed in Hansen and Scheinkman(1995) for analyzing such properties as time reversibility. But unlike their method, my approach here does not need the assumption of time reversibility. Therefore, the definition using C_0 is enough and my method is less restricted and more applicable.

$$\mathcal{A}_\theta f(x) = \sum_{i=1}^{d} b_i(x;\theta) f_i'(x) + \frac{1}{2}\sum_{i,j=1}^{d} a_{i,j}(x) f_{i,j}''(x), f \in D(\mathcal{A}), x \in \mathbb{R}^d \tag{5.5}$$

where

$$a_{ij}(x) = \sum_{k=1}^{d} \sigma_{i,k}(x)\sigma_{j,k}(x)$$

To illustrate the convenience and rich information contained about the process for the infinitesimal operator, we consider a univariate diffusion model here defined as $dX_t = b(X_t)dt + \sigma(X_t)dW_t$ with W_t a 1-dimensional standard Brownian motion in $\mathbb{R}$, $b : E \subset \mathbb{R} \to \mathbb{R}$ a drift function and $\sigma : E \to \mathbb{R}$ a diffusion function. Then by (5.5) and the discussions above, the infinitesimal operator for this univariate diffusion is

$$\mathcal{A}f(x) = b(x)f'(x) + \frac{1}{2}\sigma^2(x)f''(x) \tag{5.6}$$

Clearly the first term involving the first derivative of function $f(\cdot)$ is related to the dynamics of drift and the second term involving the second derivative of function $f(\cdot)$ to the dynamics of diffusion function. This is consistent with the intuition that drift describes the dynamics of mean and the diffusion describes that of variance of the process(see Nelson 1990 for more discussion which proves that the diffusion process is the approximation of an ARCH process). Of course, this intuition should not be taken literally due to the continuous nature of the time. Consider the infinitesimal changes of this univariate diffusion process. By (5.4) and (5.6), for any $f \in D(\mathcal{A})$, it holds P-$a.s.$ that

$$E\left(\frac{f(X_{t+h}) - f(X_t)}{h} | \mathfrak{F}_t\right) = b(X_t)f^{'}(X_t) + \frac{1}{2}\sigma^2(X_t)f^{''}(X_t) + o(h), \tag{5.7}$$

as $h \downarrow 0$. Therefore, the dynamics of $\{X_t\}$ are characterized completely by the drift and diffusion coefficients, including the conditional probability law. In contrast, for discrete time series models, the mean and variance solely cannot determine the complete conditional probability law unless it is a Gaussian process. Therefore, it is not right to simply think of drift and diffusion terms as the straightforward continuous time counterparts of conditional mean and variance respectively. In fact, the conditional mean of the process $\{X_t\}$, $E[X_{t+h}|X_t]$ for a fixed $h > 0$ is a function of both the drift $b(\cdot)$ and diffusion $\sigma(\cdot)$ instead of the drift solely(see Ait-Sahalia 1996a). The precise interpretation for drift and diffusion functions are actually (Stanton, 1997):

$$\begin{aligned} b(X_t) &= \lim_{h \to 0} E\left[\frac{X_{t+h} - X_t}{h}|X_t\right] \\ \sigma^2(X_t) &= \lim_{h \to 0} E\left[\frac{(X_{t+h} - X_t)^2}{h}|X_t\right] \end{aligned}$$

which are called instantaneous conditional mean and variance[3].

Since the diffusion process in (5.1) is a Feller process, we have three complete characterizations of the dynamics available now: transition function(or transition density), semi-group of operators and infinitesimal operator. The transition function has already been used intensively in econometric inference of diffusion models, not only in estimation (Lo 1988; Ait-Sahalia 2002; Pedersen 1995) but also in hypothesis testing (Ait-Sahalia, Fan and Peng 2008; Hong and Li 2005;). However, as we know, the transition density of most continuous time models has no closed form. Therefore, those methods based on transition density are usually computationally burdensome and inconvenient to be applied in practice. In contrast, from the discussions above, the infinitesimal operator of a diffusion process always has a closed-form and fully characterizes the dynamics.

[3]These definitions are employed by Stanton(1997) and Bandi and Phillips(2003) to propose nonparametric estimators for drift and diffusion functions.

This nice property, therefore, makes the infinitesimal operator a convenient tool for analyzing the diffusion models. It has already been used in identification (Hansen, Scheinkman and Touzi 1998), estimation (Hansen and Scheinkman 1995; Kessler and Sorenson 1999) and also hypothesis testing (Kanaya 2007; Song 2011). We shall consider the estimation problem in this study and generate convenient conditional moment restrictions by which two estimators will be proposed for the drift parameters.

To obtain moment conditions by utilizing the closed-form infinitesimal operator, we consider a transformation based on the celebrated "martingale problems". This transformation gives us a martingale characterization for diffusion processes which not only is a complete identification but also is very simple and convenient to use. By Ch5.4 of Karatzas and Shreve(1991), a probability measure P on $\left(C[0,\infty)^d, \mathcal{B}(C[0,\infty)^d)\right)$ under which

$$M_t^f = f(X_t) - f(X_0) - \int_0^t (\mathcal{A}f)(X_s)ds \tag{5.8}$$

is a martingale for every $f \in D(\mathcal{A})$, is called a solution to the martingale problem associated with the operator $\mathcal{A}$. How is the "martingale problems" related to the diffusion model? As we know, a SDE has two types of solutions: strong solutions and weak solutions(see Karatzas and Shreve(1991), Ch5.2-3 or Rogers and Williams(2000), ChV.2-3 for details). Intuitively, the strong solution is a solution to SDE with $a.s.$properties and a weak solution is that to SDE with in law properties. When the drift and diffusion terms of a SDE satisfy the Lipschitz and linear growth conditions, there is a strong solution to the SDE. But for general drift and diffusion terms, a strong solution may not exist;in this case, probabilists usually attempt to solve the SDE in the "weak" sense of finding a solution with the right probability law. The martingale problem is a variation of this "weak solution approach" developed by Strook and Varadhan(1969) and is in fact equivalent to the weak solution of a SDE. That is, the process $\{X_t\}$ is a weak solution to the SDE (2.1) if and only if

$$M_t^f(\theta) = f(X_t) - f(X_0) - \int_0^t (\mathcal{A}_\theta f)(X_s)ds \tag{5.9}$$

is a martingale for every $f \in D(\mathcal{A})$, where $\mathcal{A}_\theta$ is defined as in (5.5). For detailed discussions and proof, see ChV.19-20 of Rogers and Williams(2000), Theorem 21.7 of Kallengberg(2002), or Proposition 2.4 of ChVII in Revuz and Yor(2005). One point worth mentioning is that when strong solution exists the weak solution will coincide with it. Hence it is enough to consider the weak solution identification for doing econometric inference because regularity conditions for the existence of strong solution are usually satisfied and thus imposed in analysis. See Protter(2005) for some regularity Lipschitz conditions for the existence and uniqueness of a strong solution to a SDE.

Now we have shown that the identification of the multivariate time-homogeneous diffusion model in (5.1) is equivalent to the martingale property of the transformed processes in (5.9), which can be written as a conditional mean restriction:

$$E\left[M_t^f(\theta)|\mathcal{I}_{t'}\right] = M_{t'}^f(\theta)$$

for any $f \in D(\mathcal{A})$ and $t' < t$, where call$\mathrm{I}_{t'} = \sigma\{X_{t''}\}_{t''<t'}$ is the sigma-field generated by the past information of $\{X_t\}$ at time t'. It is well known that a GMM estimator can be derived based on the conditional moment restriction. However, observe that here we have infinite many conditional moment restrictions because there are usually an infinite number of functions $f(\cdot)$ in the domain $D(\mathcal{A})$ which are usually called test functions. It is very difficult and burdensome in

practice, although maybe not impossible, to construct a GMM estimator based on these infinitely many conditional moment conditions due to the difficulty of exhausting all possible function forms of $f(\cdot)$ in $D(\mathcal{A})$. This is a general problem which appears not only in my study here but also for all the other papers employing infinitesimal operators, like Hansen and Scheinkman(1995), Conley, Hansen, Luttmer and Scheinkman (1997), and Kanaya (2007). To tackle such a difficulty, the space of test functions has to be reduced to an equivalent subclass. Kanaya (2007) does this based on the concept of a core and "approximation" theory. Hansen and Scheinkman (1995) and Conley, Hansen, Luttmer and Scheinkman (1997) also discuss the choice of test functions. But no formal evidence is provided for the equivalence and no loss of information and identification. In contrast, based on a celebrated theorem in probability theory, a subclass of $D(\mathcal{A})$ which not only consists of finitely many function forms but also plays the same role as $D(\mathcal{A})$ is obtained for the martingale characterization (5.9). By Proposition 4.6 and Remark 4.12 of Karatzas and Shreve (1991, Ch5.4), the process $\{X_t\}$ is a weak solution to the SDE in (5.1) if it satisfies the martingale problem with $\mathcal{A}$ as the infinitesimal operator of $\{X_t\}$ for the choices $f(x) = x_i$ and $f(x) = x_i x_j$ with $1 \leq i, j \leq d$. Therefore, the process $\{X_t\}$ is a weak solution to the SDE (5.1) if

$$\begin{aligned}
M_t^{x_i}(\theta_0) &= X_t^i - X_0^i - \int_0^t b_i(X_s;\theta_0)ds \\
M_t^{x_i,x_i}(\theta_0) &= \left(X_t^i\right)^2 - \left(X_0^i\right)^2 - \int_0^t \left[2b_i(X_s;\theta_0)X_s^i + \sum_{k=1}^d \sigma_{i,k}(X_s)^2\right] ds \\
M_t^{x_i,x_j}(\theta_0) &= X_t^i X_t^j - X_0^i X_0^j \\
&\quad - \int_0^t \left[b_i(X_s;\theta_0)X_s^j + b_j(X_s;\theta_0)X_s^i + \frac{1}{2}\sum_{k=1}^d \sigma_{i,k}(X_s)\sigma_{j,k}(X_s)\right] ds
\end{aligned} \tag{5.10}$$

with $i \neq j$, are martingales for $1 \leq i, j \leq d$. Of course, the converse of this result only holds with local martingale replacing martingale. But since examples which are local martingales but not martingales are few and too artificial in certain sense even when they exist[4], I regard them as almost the same and do not pay much attention to their difference in this study[5]. Henceforth, the identification of the diffusion model in (5.1) is transformed into the martingale property of the transformed processes.

The resulting characterization (5.10) is much more simple and intuitive than those in Kanaya (2007), Hansen and Scheinkman (1995), and so on. It greatly simplifies these conditional moment restrictions and makes the derivation of a GMM estimator based on them completely practical. Two points are worth noting here. The first one is that the conditional moment restrictions can be expressed explicitly by the drift and diffusion terms. Therefore, they can be used directly while in contrast, the transition density based methods like Lo (1988) and Ait-Sahalia (2002) have to either approximate the transition density or numerically solve it because the transition density rarely has a closed-form. The second is that the identification of a multivariate d-dimensional diffusion process is equivalent to the martingale property for $d' = (d^2 + 3d)/2$ univariate processes which are explicit expressions of drift and diffusion terms. This makes the conditional moment restrictions particularly convenient for multivariate diffusion models for which the transition density methods are extremely complicated and computationally inconvenient.

To have a more intuitive understanding of the characterization (5.10), we

[4]See Karatzas and Shreve(1991), p.168 and 200-201 for some examples which are local martingales but not martingales.

[5]When the difference really matters, the local martingale property can be explored by the time-change techniques. The idea is to use the fact that the time-changed continuous local martingale by quadratic variation is a standard Brownian Motion(see Andersen, Bollerslev & Dobrev(2007) and Park(2008) for details). Since this approach is closely related to time-dependent diffusion models and the estimation will be very different, I do not pursue it here.

consider the simplified version for univariate diffusion models:

$$
\begin{aligned}
M_t^x(\theta_0) &= X_t - X_0 - \int_0^t b(X_s;\theta_0)ds \\
M_t^{x^2}(\theta_0) &= X_t^2 - X_0^2 - \int_0^t \left[2b(X_s;\theta_0)X_s + \sigma^2(X_s)\right] ds
\end{aligned}
\tag{5.11}
$$

are both martingales. Observe that the first transformed process M_t^x only involves the drift term and the second $M_t^{x^2}$ has both the drift and diffusion terms as inputs. Intuitively, M_t^x characterizes the dynamics of the drift term solely and this characterization is robust to the dynamics of diffusion term. Note also that $\int_0^t \sigma^2(X_s)ds$ is the so-called "integrated volatility" or the quadratic variation $[X, X]_t$ of the process $\{X_t\}$ which has received extremely intensive attention in recent years(see Andersen, Bollerslev, Diebold, and Labys, 2003; Barndorff-Nielsen and Shephard, 2006; Ait-Sahalia, Mykland and Zhang, 2005). Therefore $M_t^{x^2}$ contains the dynamics of diffusion term, i.e., the volatility of the process illustrated by $\int_0^t \sigma^2(X_s)ds$. Furthermore, $M_t^{x^2}$ also characterizes the interaction between drift and diffusion terms which is represented by $\int_0^t b(X_s;\theta_0)X_s ds$ because $b(X_s;\theta_0)X_s$ will raise the power of X_s at least to 2 and hence variance will also appear in this term.[6]

Now for the convenience of constructing estimators, we state the following conditional moment restrictions using the *m.d.s.*property for the first-order difference of the transformed processes. For the multivariate diffusion model in (2.1), the identification is equivalent to the following conditional moment restriction:

$$
E\left[Z_t(\theta_0)|\mathcal{I}_{t'}\right] = 0
$$

[6]The characterization (2.11) is essentially equivalent to the celebrated Levy Characterization of Brownian Motion(Øksendal, 2003, Theorem 8.6.1) if we take $b(\cdot) \equiv 0$ and $\sigma(\cdot) \equiv 1$. Hence, the infinitesimal operator based characterization is actually an extension of this Levy Characterization Theorem to general diffusion models. See Chapter 5 for more detailed discussions.

for any $t' < t$, where $\text{callI}_{t'} = \sigma\{X_{t''}\}_{t''<t'}$ is the sigma-field generated by the past information of $\{X_t\}$ at time t' and $Z_t(\theta)$ is a vector with components for $i, j = 1, \cdots, d$

$$\begin{aligned}
Z_t^i(\theta_0) &= M_t^{x_i}(\theta_0) - M_{t-\Delta}^{x_i}(\theta_0) \\
&= X_t^i - X_{t-\Delta}^i - \int_{t-\Delta}^t b_i(X_s;\theta_0)ds \qquad (5.12)\\
Z_t^{i,i}(\theta_0) &= M_t^{x_i x_i}(\theta_0) - M_{t-\Delta}^{x_i x_i}(\theta_0) \\
&= \left(X_t^i\right)^2 - \left(X_{t-\Delta}^i\right)^2 - \int_{t-\Delta}^t \left[2b_i(X_s;\theta_0)X_s^i + \sum_{k=1}^d \sigma_{i,k}(X_s)^2\right]ds \\
Z_t^{i,j}(\theta_0) &= M_t^{x_i x_j}(\theta_0) - M_{t-\Delta}^{x_i x_j}(\theta_0) \\
&= X_t^i X_t^j - X_{t-\Delta}^i X_{t-\Delta}^j - \int_{t-\Delta}^t \Big[b_i(X_s;\theta_0)X_s^j + b_j(X_s;\theta_0)X_s^i \\
&\quad + \frac{1}{2}\sum_{k=1}^d \sigma_{i,k}(X_s)\sigma_{j,k}(X_s)\Big]ds \qquad (5.13)
\end{aligned}$$

for $i \neq j$.

Corresponding to (5.12) and (5.13), the identification of univariate diffusion models is equivalent to the following conditional moment restrictions:

$$E\left[Z_t(\theta_0)|\mathcal{I}_{t'}\right] = 0$$

for any $t' < t$, where $Z_t(\theta_0) = \left(Z_t^x(\theta_0), Z_t^{x^2}(\theta_0)\right)'$, $\text{callI}_{t'} = \sigma\{X_{t''}\}_{t''<t'}$, and

$$\begin{aligned}
Z_t^x(\theta_0) &= M_t^x(\theta_0) - M_{t-\Delta}^x(\theta_0) \\
&= X_t - X_{t-\Delta} - \int_{t-\Delta}^t b(X_s;\theta_0)ds \qquad (5.14)\\
Z_t^{x^2}(\theta_0) &= M_t^{x^2}(\theta_0) - M_{t-\Delta}^{x^2}(\theta_0) \\
&= X_t^2 - X_{t-\Delta}^2 - \int_{t-\Delta}^t \left[2b(X_s;\theta_0)X_s + \sigma^2(X_s)\right]ds \qquad (5.15)
\end{aligned}$$

5.2 The First Estimator: Integrating Diffusion Functions via Quadratic Variation and Covariation

In this section, we shall construct the first estimator based on the conditional moment restrictions in (5.12) and (5.13) for the multivariate semi-parametric diffusion models in (5.1). As an illustration, the estimator for univariate diffusion models is also presented as a special case. The sample data is discrete in time, i.e., $\{X_{\tau\Delta}\}_{\tau=1}^{n}$ observed over a time span T with sampling interval Δ and sample size $n = T/\Delta$. Therefore, the process is in continuous time but the data sample is discrete. This is a general problem in continuous-time series econometrics(see Lo (1988) and Ait-Sahalia (1996a,b) for discussions about the estimation of the discretized version of a continuous-time model). The asymptotic schemes we employ for the first estimator are $n = T/\Delta \to \infty$ and for each sampling interval Δ, we have high frequency data with the sampling interval $h = \Delta/m \to 0$ for integer m. The former is a standard treatment in the literature of estimating diffusion models (Ait-Sahalia, 2002; Hansen and Scheinkman, 1995) while the latter is here to ensure the consistency of realized volatility(covariation) to the quadratic variation (covariation) by which we integrate out the unknown diffusion function.

The conditional moment restriction is $E\left[Z_t(\theta)|\mathcal{I}_{t'}\right] = 0$ for any $t' < t$, where call$\mathrm{I}_{t'}$=$\sigma\{X_{t''}\}_{t''<t'}$ is the sigma-field generated by the past information of $\{X_t\}$ and $Z_t(\theta_0)$ is a vector with components defined in (5.12) and (5.13). An application of the Law of Iterated expectation implies that $E\left[Z_{\tau\Delta}(\theta)|\mathcal{I}_{\tau-1}\right] = 0$, where

$$\mathcal{I}_{\tau-1} = \sigma\{X_{(\tau-1)\Delta}, X_{(\tau-2)\Delta}, \cdots, X_{\Delta}\} \tag{5.16}$$

Observe that (5.16) is a *m.d.s.* property for discrete time process $\{Z_{\tau\Delta}(\theta)\}_{\tau=1}^{n}$ and it is derived as an implication of the *m.d.s.*property in continuous time instead of a result from the discretization of the continuous time process. In this respect, it is similar to the approaches of Ait-Sahalia (1996a,b) and Lo (1988) and therefore is free of the discretization errors which are discussed in Lo (1988).

By the Markov property,

$$E\left[Z_{\tau\Delta}(\theta)|X_{(\tau-1)\Delta}\right] = 0 \tag{5.17}$$

for any $\tau \geq 1$. This will be the conditional moment condition we will depend on for proposing the estimator and we assume (5.17) holds for a unique value $\theta_0 \in \Theta$. Therefore, the problem we have now is to estimate θ_0, the $q \times 1$ vector of unknown parameters, in the following conditional restriction:

$$E\left[Z(\theta_0)|X\right] = 0 a.s. \tag{5.18}$$

using sample data $\{Z_{\tau\Delta}\}_{\tau=2}^{n}$, where Z is a $d' \times 1$ vector with components defined as in (5.13) and X is a $d \times 1$ vector of conditioning variables. Without loss of generality, I assume X is bounded with probability one; see e.g., Bierens (1994).

It is well known that (5.18) is equivalent to the unconditional moment restrictions:

$$E\left[Z(\theta_0)g(X)\right] = 0 \tag{5.19}$$

for all measurable functions g, where each $g(Z)$ may be interpreted as an "instrument" that helps to identify θ_0. In practice, it is infeasible to consider all possible functions. Hence one typical method is to form an estimating eqnarray by choosing certain instruments subjectively, such as the square and cross product of the elements in X. Then some suitable estimation methods can be applied to obtain the parameter estimators, such as the GMM of Hansen (1982) or the empirical likelihood method of Qin and Lawless (1994) and Kitamura (1997). There is no problem for this approach in a linear model because any subset of linearly independent unconditional moment restrictions identifies θ_0 globally as long as the dimension of this subset equals that of θ_0. However, from (5.13), $Z(\theta_0)$ is generally nonlinear in θ_0. And in this case, as shown by Domínguez and Lobato (2004), θ_0 is not necessarily identified globally when unconditional moments are chosen arbitrarily and the identification may depend on the marginal distributions of the conditioning variables X.

To be free of the identification problem, we shall follow Chapter 5 to explore the conditional moment restriction directly, based on a special choice of g in (5.19) which makes the unconditional moment restriction (5.19) equivalent to the conditional moment restriction (5.18). Specifically, we take the indicator functions as the instruments. By Billingsley (1995, Theorem 16.10iii),

$$E\left[Z(\theta_0)|X\right] = 0 a.s. \iff H\left(\theta_0, x\right) = 0 \tag{5.20}$$

for almost all $x \in \mathbb{R}^d$, where $H(\theta, x) \equiv E\left[Z(\theta) 1(X \le x)\right]$ and the indicator function $1(X \le x) \equiv {}_{m=1}^{d} 1(X_m \le x_m)$. Since (5.17) holds for a unique value $\theta_0 \in \Theta$, it follows that $P(E[Z(\theta)|X] = 0) < 1$ when $\theta \neq \theta_0$ and hence $H(\theta_0, x) \neq 0$ in a non-null set of the sample space of X. Therefore, denoting by $P_{X_{\tau-1}}$ the probability distribution functions of the random vector $X_{\tau-1}$, $\int |H(\theta_0, x)|^2 \, dP_{X_{\tau-1}}(x) = 0$ but $\int |H(\theta, x)|^2 \, dP_{X_{\tau-1}}(x) \neq 0$ for any $\theta \neq \theta_0$. Then θ_0 can be globally identified this integral, i.e.,

$$\theta_0 = \arg\min_{\theta \in \Theta} \int |H(\theta, x)|^2 \, dP_{X_{\tau-1}}(x) \tag{5.21}$$

and θ_0 is the unique value that satisfies (5.21). We first construct the estimator by assuming $\sigma^2(\cdot)$ is known and only θ_0 has to be taken care of and then deal with the unknown $\sigma^2(\cdot)$ via quadratic variation(covariation). Denote the sample analog for $H(\theta, x)$ by $H_{n-1}(\theta, x) = (n-1)^{-1}\sum_{\tau=2}^{n} Z_\tau(\theta) 1(X_{\tau-1} \le x)$. Similar to Domínguez and Lobato (2004), we can estimate θ_0 by the sample analog of (5.21), that is

$$
\begin{aligned}
\widehat{\theta_0} &= \arg\min_{\theta\in\Theta} \frac{1}{n-1}\sum_{l=2}^{n} |H_{n-1}(\theta, X_l)|^2 \\
&= \arg\min_{\theta\in\Theta} \frac{1}{n-1}\sum_{l=2}^{n} \left|\frac{1}{n-1}\sum_{\tau=2}^{n} Z_\tau(\theta) 1(X_{\tau-1} \le X_l)\right|^2 \\
&= \arg\min_{\theta\in\Theta} \sum_{a=i,(i,j);i,j=1,\cdots,d} \left\{\frac{1}{n-1}\sum_{l=2}^{n}\left(\frac{1}{n-1}\sum_{\tau=2}^{n} Z_\tau^a(\theta) 1(X_{\tau-1} \le X_l)\right)^2\right\} \qquad (5.22)
\end{aligned}
$$

This is a minimum distance estimator. It does not involve either matrix inversion or nonparametric estimation and thus computationally convenient although an additional summation of $n-1$ terms needs be computed compared to the standard GMM objective function. Actually, for most parametric specifications of the drift functions, the estimator has an analytic formula.

As noted earlier, (5.22) involves the unknown diffusion function $\sigma(\cdot)$. To to make the estimator $\widehat{\theta_0}$ robust to the mis-specification of the diffusion term, $\sigma(\cdot)$ has to be dealt with nonparametrically. One potential approach is the nonparametrically smoothing method. For example, the kernel method can be used to estimate the diffusion function due to its simplicity and intuitive appeal. Then the sample analog of (5.22) can be formed replacing $\sigma(\cdot)$ by its kernel estimator $\widehat{\sigma}(\cdot)$. But this approach will affect the asymptotic property of the estimator of θ_0 because it is well known that the nonparametric estimation has a slower convergence rate than that of parametric estimation. If we replace $\sigma(\cdot)$ by an estimator

$\widehat{\sigma}(\cdot)$ with higher convergence rate than $\sqrt{n}$, the asymptotic theory is still determined by the parametric estimation of θ_0. But if the convergence rate of $\widehat{\sigma}(\cdot)$ is less than or just equal to $\sqrt{n}$, the asymptotic property of $\widehat{\theta_0}$ will be affected by the estimation error of $\widehat{\sigma}(\cdot)$ and hence nonstandard. This makes the econometric inference difficult and inconvenient. In addition, the nonparametric method introduces a bandwidth number to which we have to analyze the sensitivity of the estimator. The procedure then becomes more inconvenient in practice. It is actually very similar to that of Kristensen (2008a) with the difference that the latter uses the more complicated simulation based transition density and the former does not involve transition density at all. Although this approach is already very easier to apply in practice than that of Kristensen (2008a), it still has restricted applicability since due to the "curse of dimensionality", it is extremely unsuitable for multivariate models.

To obtain a more convenient and easier-to-implement estimator, we choose to employ an approach based on quadratic variation(covariation). To see the relationship between our infinitesimal operator based conditional moment restriction and quadratic variation(covariation), we first consider the moment conditions in (5.14) and (5.15) for univariate models. It can be observed from (5.14) and (5.15) that the diffusion function only appears in $Z_\tau^{x^2}(\theta)$ with the form $\int_0^t \sigma^2(X_s)ds$. It is well known that for the diffusion model, $\int_0^t \sigma^2(X_s)ds$ is equal to the quadratic variation $[X,X]_t$(also known as integrated volatility) which has been analyzed intensively in recent years (see Andersen, Bollerslev, Diebold, and Labys, 2003; Barndorff-Nielsen and Shephard 2002, 2006; Ait-Sahalia, Mykland and Zhang, 2005). Now change the notations and then we have

$$Z_t^{x^2}(\theta_0) = Z_t^{x^2}(\theta_0, [X,X])$$

$$= \quad X_t^2 - X_{t-\Delta}^2 - \int_{t-\Delta}^{t} 2b(X_s;\theta_0)X_s ds - [X,X]_{t-\Delta}^t \tag{5.23}$$

where $[X,X]_{t-\Delta}^t = [X,X]_t - [X,X]_{t-\Delta}$. Therefore the diffusion function is integrated out by the quadratic variation which can then be estimated consistently by the so-called realized volatility:

$$\widehat{[X,X]}_{t-\Delta}^t = \sum_{i=1}^{m} \left(X_{t-\Delta+ih} - X_{t-\Delta+(i-1)h}\right)^2 \tag{5.24}$$

Of course, here the infill asymptotic scheme has to be assumed, i.e., $h \to 0$. Consequently, $m \to \infty$ if Δ is fixed and

$$\widehat{[X,X]}_{t-\Delta}^t = O_p(h^{1/2}) = O_p\left(\sqrt{\Delta/m}\right) \tag{5.25}$$

by Barndorff-Nielsen and Shephard (2002) and Bandi and Russell (2008).

This approach of integrating out diffusion functions by quadratic variation has much better properties than that of smoothing diffusion function by nonparametric methods. First, the estimator for the quadratic variation, i.e., the realized volatility has a much higher convergence rate than the nonparametric smoothing estimator for diffusion function. Second, the realized volatility is essentially nonparametric and does not involve the choice of any other parameters. In contrast, one has to choose the kernel function and smoothing bandwidth in nonparametric smoothing methods and the choice of the latter is usually difficult and no universal standard exists. Third, the convergence rate of the realized volatility to quadratic variation is $(m/\Delta)^{1/2}$ while for nonparametric smoothing estimator to converge to the true diffusion function, the convergence rate is usually low especially for the multivariate case due to the "curse of dimensionality" and the finite sample performance may not be reliable. Fourth, the quadratic variation method can be extended easily to multivariate case using the so-called quadratic covariation. It is free of the "curse of dimensionality" which is generally suffered by nonparametric smoothing methods.

As discussed above, the pre-estimated processes $\widehat{Z_t} \equiv Z_t(\theta, \widehat{[X, X]}) \equiv \left(Z_t^x(\theta), Z_t^{x^2}(\theta, \widehat{[X, X]})\right)$ can be obtained. Denote the sample analog for $H(\theta, x) = H(\theta, x, [X, X])$ by

$$H_{n-1}\left(\theta, x, \widehat{[X, X]}\right) = (n-1)^{-1} \sum_{\tau=2}^{n} Z_{\tau\Delta}(\theta, \widehat{[X, X]}) 1 \left(X_{\tau-1} \leq x\right) \tag{5.26}$$

The realized volatility based estimator $\widehat{\theta}_{RV}$ can be obtained by

$$\begin{aligned}
\widehat{\theta}_{RV} &= \arg\min_{\theta} \frac{1}{n-1} \sum_{l=2}^{n} \left| H_{n-1}\left(\theta, x, \widehat{[X, X]}\right) \right|^2 \\
&= \arg\min_{\theta} \frac{1}{n-1} \sum_{l=2}^{n} \left| \frac{1}{n-1} \sum_{\tau=2}^{n} Z_{\tau}(\theta, \widehat{[X, X]}) 1 \left(X_{\tau-1} \leq X_l\right) \right|^2 \\
&= \arg\min_{\theta} \frac{1}{(n-1)^3} \sum_{l=2}^{n} \left\{ \left(\sum_{\tau=2}^{n} Z_{\tau}^{x}(\theta) 1 \left(X_{\tau-1} \leq X_l\right) \right)^2 \right. \\
&\quad \left. + \left(\sum_{\tau=2}^{n} Z_{\tau}^{x^2}(\theta, \widehat{[X, X]}) 1 \left(X_{\tau-1} \leq X_l\right) \right)^2 \right\}
\end{aligned} \tag{5.27}$$

Similarly for the general multivariate semi-parametric diffusion models, we can observe from (5.12) and (5.13) that the diffusion functions can be integrated out by the so-called "quadratic covariation" $\left[X_i, X_j\right]_t$ which is equal to $\int_0^t \sum_{k=1}^{d} \sigma_{i,k}(X_s)\sigma_{j,k}(X_s) ds$ and which has been analyzed by Barndorff-Nielsen and Shephard (2004a), Bandi and Russell (2005) and Zhang (2006). We denote

$$\begin{aligned}
Z_t^{i,i}(\theta_0) &= Z_t^{i,i}(\theta_0; [X_i, X_i]) \\
&= \left(X_t^i\right)^2 - \left(X_{t-\Delta}^i\right)^2 \\
&\quad - \int_{t-\Delta}^{t} \left[2 b_i(X_s; \theta_0) X_s^i\right] ds - \left[X^i, X^i\right]_{t-\Delta}^{t} \\
Z_t^{i,j}(\theta_0) &= Z_t^{i,j}(\theta_0; \left[X_i, X_j\right]) \\
&= X_t^i X_t^j - X_{t-\Delta}^i X_{t-\Delta}^j - \int_{t-\Delta}^{t} \left[b_i(X_s; \theta_0) X_s^j + b_j(X_s; \theta_0) X_s^i\right] ds \\
&\quad - \frac{1}{2} \left[X^i, X^j\right]_{t-\Delta}^{t}
\end{aligned} \tag{5.28}$$

for $i \neq j$, where $\left[X_i, X_j\right]_{t-\Delta}^{t} = \left[X_i, X_j\right]_t - \left[X_i, X_j\right]_{t-\Delta}$. Therefore the diffusion functions are all integrated out by the quadratic variation and covariation. The following so-called realized covariation can be used which is a consistent estimator for the quadratic covariation:

$$\widehat{[X^i, X^j]}_{t-\Delta}^{t} = \sum_{i=1}^{m} \left(X_{t-\Delta+ih}^{i} - X_{t-\Delta+(i-1)h}^{i} \right) \left(X_{t-\Delta+ih}^{j} - X_{t-\Delta+(i-1)h}^{j} \right) \tag{5.29}$$

By Barndorff-Nielsen and Shephard (2004) and similar to (5.25), we have

$$\widehat{[X^i, X^j]}_{t-\Delta}^{t} = O_p(h^{1/2}) = O_p\left(\sqrt{\Delta/m}\right) \tag{5.30}$$

as long as $h \to 0$. Let $[X, X]$ denote the matrix of quadratic variation and covariation and $\widehat{[X, X]}$ denote the corresponding matrix of estimators. Then the pre-estimated vector process $\widehat{Z}_t \equiv Z_t(\theta, \widehat{[X, X]})$ has components $Z_t^i(\theta)$, $Z_t^{i,i}(\theta, \widehat{[X_i, X_i]})$, and $Z_t^{i,j}(\theta, \widehat{\left[X_i, X_j\right]})$. For notation simplicity, I also use $\widehat{Z}_{\tau,a}$ for $a = i, ij$ with $i, j = 1, \cdots, d$ to denote the components of $\widehat{Z}_\tau$ sometimes when there is no confusion. Now I can estimate the multivariate version of generalized As discussed above, the pre-estimated processes $\widehat{Z}_t \equiv Z_t(\theta, \widehat{[X, X]}) \equiv \left(Z_t^x(\theta), Z_t^{x^2}(\theta, \widehat{[X, X]})\right)$ can be obtained. Denote the sample analog for $H(\theta, x) = H(\theta, x, [X, X])$ by

$$H_{n-1}\left(\theta, x, \widehat{[X, X]}\right) = (n-1)^{-1} \sum_{\tau=2}^{n} Z_{\tau\Delta}(\theta, \widehat{[X, X]}) 1\left(X_{\tau-1} \leq x\right) \tag{5.31}$$

The realized volatility based estimator $\widehat{\theta}_{RV}$ for the general multivariate semi-parametric diffusion model can be obtained by

$$\begin{aligned}
&\widehat{\theta}_{RV} \\
= \;& \arg\min_{\theta} \frac{1}{n-1} \sum_{l=2}^{n} \left| H_{n-1}\left(\theta, x, \widehat{[X, X]}\right) \right|^2 \\
= \;& \arg\min_{\theta} \frac{1}{n-1} \sum_{l=2}^{n} \left| \frac{1}{n-1} \sum_{\tau=2}^{n} Z_\tau(\theta, \widehat{[X, X]}) 1\left(X_{\tau-1} \leq X_l\right) \right|^2 \\
= \;& \arg\min_{\theta\in\Theta} \frac{1}{(n-1)^3} \sum_{a=i,(i,j);i,j=1,\cdots,d} \\
& \left\{ \frac{1}{n-1} \sum_{l=2}^{n} \left(\frac{1}{n-1} \sum_{\tau=2}^{n} Z_\tau^a(\theta) 1\left(X_{\tau-1} \leq X_l\right) \right)^2 \right\}
\end{aligned} \tag{5.32}$$

for which we shall prove the asymptotic properties in Section 5.4.

5.3 The Second Estimator: Exploring Separate Identification

In this section, the second estimator for drift parameters in the general multivariate semi-parametric diffusion model will be proposed via a special property of the infinitesimal operator based characterization, i.e., the infinitesimal operator of the diffusion process is a closed-form expression of drift and diffusion terms in an essentially separate manner. Such a nice property is enjoyed neither by the stationary density or by the transition density. Consequently, the resulting conditional moment restrictions can also identify the dynamics of the drift and diffusion functions separately, which enables us to estimate the drift parameters robust to the diffusion function mis-specification. Suppose the sample data we have are $\{X_{\tau\Delta}\}_{\tau=1}^{n}$ observed over a time span T with sampling interval Δ and sample size $n = T/\Delta$. Different from those for the first estimator, the asymptotic scheme we employ in this section estimator is $n = T/\Delta \to \infty$. It can be obtained by either infill ($\Delta \to 0$) or long span ($T \to \infty$) instead of both and this implies that our second estimator can be applied to both high-frequency and low-frequency data.

We first consider the univariate models for simple illustration. As noticed earlier, it can be seen from (5.11) that the first transformed process M_t^x involves only the drift term and the second $M_t^{x^2}$ has both the drift and diffusion terms as inputs. Intuitively, M_t^x characterizes the dynamics of the drift term solely and this characterization is robust to the dynamics of diffusion term. Therefore, we can estimate the drift parameter θ_0 solely robust to diffusion mis-specification if we further assume the drift parameters are identified by the condition that M_t^x is a martingale. That is, there exists a unique $\theta_0 \in \Theta$ such that

$$M_t^x(\theta_0) = X_t - X_0 - \int_0^t b(X_s; \theta_0) ds \tag{5.33}$$

is a martingale. Following the same reasoning as that in last section, the identification of the drift parameter θ_0 can be written as the following conditional moment restriction:

$$E\left[Z_t(\theta_0)|X_{\tau-1}\right] = 0 a.s.$$

for $\tau \geq 1$, where

$$Z_t(\theta_0) = M_t^x(\theta_0) - M_{t-\Delta}^x(\theta_0) = X_t - X_{t-\Delta} - \int_{t-\Delta}^t b(X_s; \theta_0) ds \tag{5.34}$$

Then similar to (5.21), θ_0 can be globally identified as follows:

$$\theta_0 = \arg\min_\theta \int \left[H(\theta, x)\right]^2 dP_{X_{\tau-1}}(x) \tag{5.35}$$

where $H(\theta, x) \equiv E\left[Z_\tau(\theta) 1 \left(X_{\tau-1} \leq x\right)\right]$ and θ_0 is the unique value that satisfies (5.35). Denote the sample analog for $H(\theta, x)$ by $H_{n-1}(\theta, x) = (n-1)^{-1} \sum_{\tau=2}^n Z_\tau(\theta) 1 \left(X_{\tau-1} \leq x\right)$. Similar to (5.22), our second estimator for the drift parameter θ_0 via the separate identification condition can be obtained as:

$$\widehat{\theta}_{SI} = \arg\min_{\theta\in\Theta} \frac{1}{n-1} \sum_{l=2}^n \left[\frac{1}{n-1} \sum_{\tau=2}^n Z_\tau(\theta) 1 \left(X_{\tau-1} \leq X_l\right)\right]^2 \tag{5.36}$$

where the subscript SI stands for separate identification. This is also a minimum distance estimator. Note that similar to Ait-Sahalia (1996b, 2002), Hong and Li 2005) and Song (2011), our second estimator only depends on discrete time implications of the conditional moment restrictions which are derived in continuous time. As a result, it does not depend on the sampling interval Δ and hence its asymptotic property only requires the number of observations n to go to infinity. Therefore, it is applicable to both high and low frequency data.

Most recently, Park (2008) proposes a "conditional mean model of instantaneous change for a given stochastic process" and the identification of the model is equivalent to a martingale property in continuous time, which coincides with the separate identification condition (5.33) above. Specifically, Park's(2008) model is defined by a so-called general continuous time regression:

$$dX_t = b(X_t, \theta)dt + dU_t, \tag{5.37}$$

where $\{X_t\}$ is a stochastic process, $\{\mathcal{F}_t\}$ is a filtration to which $\{X_t\}$ is adapted, and $\{U_t\}$ is a martingale process with respect to the filtration $\{\mathcal{F}_t\}$ so that dU_t is a *m.d.s.* with $E(dU_t|\mathcal{F}_t) = 0$. The process $\{X_t\}$ can be either stationary or nonstationary as long as the drift function $b(X_t)$ is an adapted process of finite variation. The drift function $b(Y_t)$ can be interpreted as an instantaneous conditional mean, namely,

$$b(X_t) = \lim_{\Delta \to 0} E\left[\frac{X_t - X_{t+\Delta}}{\Delta} \middle| \mathcal{F}_t \right]$$

for all $0 < t < \infty$ similar to the interpretation of the drift function for a diffusion model. Following Park (2008), (5.37) is called a martingale regression in continuous-time, and dU_t is a continuous-time regression error. Eqnarray (5.37) can be viewed as the decomposition of a physical system into a signal (the finite variation process) and a noise (the local martingale).

Integrating the process in (5.37), we obtain

$$U_t = (X_t - X_0) - \int_0^t b(X_s; \theta)ds, \tag{5.38}$$

which is a martingale. One important implication is that $\varepsilon_{t,\Delta} \equiv U_t - U_{t-\Delta}$ is a *m.d.s.* for any given $\Delta > 0$. The process $\varepsilon_{t,\Delta}$ may be viewed as the true discrete-time regression error. Note that there is no discretization bias for the discrete-time regression error $\varepsilon_{t,\Delta}$ because no discretization is made. As can be seen obviously,

U_t coincides with our first transformed process M_t^x from the infinitesimal operator based characterization and $\varepsilon_{t,\Delta}$ is exactly the same as Z_t in (5.34). Therefore, our separate identification based estimator $\widehat{\theta}_{SI}$ is also applicable for the instantaneous conditional model in (5.37), which actually covers more models than just diffusion processes as special cases.

When $\{U_t\}$ has *a.s.* continuous sample paths, we can write $dU_t = \sigma_t dW_t$, where $\{\sigma_t\}$ is adapted to $\{\mathcal{F}_t\}$ and $\{W_t\}$ is the standard Brownian motion with respect to $\{\mathcal{F}_t\}$. Clearly, (5.37) includes the class of one-factor semi-parametric diffusion processes as a special case[7], where X_t is exactly the same as the univariate version of the model we consider in (5.1). The specification of $\sigma(\cdot)$ is totally unrestricted under (5.37)[8].

The error process U_t in (5.37) has a more general structure than the class of one-factor semi-parametric diffusion models in (5.1). For example, stochastic volatility (SV) (see, e.g., Anderson and Lund (1996)) is actually also incorporated as a special case due to the unrestricted specification of the volatility structure. Furthermore, while the 'normal' vibrations or smooth variation in the changes in X_t are modeled by a standard Brownian motion, the 'abnormal' vibrations may be due to the arrival of important information that has more than a marginal effect on changes in X_t. Such important information usually arrives only at discrete points in time, so this jump component is most appropriately modeled with a counting process reflecting the non-marginal impact of the in-

[7]On the other hand, Park's(2008) instantaneous conditional mean model can also be regarded as a special case of the infinitesimal operator based martingale characterization since it can be derived using a special choice of test functions(see Song(2011, Theorem 2) or discussions in the last section). If interesting function forms other than $f(x) = x_i$ and $x_i x_j$ are suitably chosen, other characterizations of diffusion processes may be obtained.

[8]Examples of $\sigma(\cdot)$ include the interest rate models of Vasicek (1977), Cox, Ingersoll and Ross(1985), and Chan *et al.* (1992). The main difference among these models lies in their functional forms for $b(\cdot)$ and $\sigma(\cdot)$ in (5.1). See, e.g., Aït-Sahalia (1996a, Table I) for alternative specifications of the spot interest rate.

formation. In (5.37), the possibility of a jump component is also embedded since $\{U_t\}$ is a general martingale process, and does not require the continuity of the sample path. In other words, we allow as a special case that U_t consists of a diffusive part and a jump component, for example, $dU_t = \sigma_t dW_t + J_t dN_t$, where J_t is the jump size, and $\{N_t\}$ is a (homogeneous) Poisson process with arrival rate λ, independent of $\{W_t\}$. Since a Poisson process N_t is a counting process increasing with time and cannot be a martingale, with the presence of jumps, U_t may not be a martingale, thus dU_t is not a *m.d.s.* In such cases, we can transform it into a martingale by subtracting a proper "mean". By Protter (2005), a compensated Poisson process, $N_t^* = N_t - \lambda t$, constructed as such will be a martingale. Thus, we can consider a continuous-time regression with a compensated Poisson (jump) process as follows:

$$
\begin{aligned}
dX_t &= [b(X_t) - \lambda\kappa]\, dt + \sigma(X_t) dW_t + J_t dN_t \\
&= \mu(X_t) dt + \sigma(X_t) dW_t + J_t dN_t^*,
\end{aligned}
$$

where $\kappa = E(J_t)$ and $N_t^* = N_t - \lambda t$, a compensated Poisson process. Thus, we will have $E(dU_t|\mathcal{F}_t) = 0$ and $U_t \equiv \sigma(Y_t) dW_t + J_t dN_t^*$. Henceforth, the jump diffusion models are also covered by (5.37) as special cases, enlarging the applicability greatly of our estimator $\widehat{\theta}_{SI}$.

Park (2008) also proposes a minimum distance estimator based on the same martingale property. But his estimator is constructed by utilizing the fact that the time-changed continuous martingale is a standard Brownian Motion. Since the time change requires estimating the quadratic variation, his estimator needs $\Delta \to 0$ and only applies to the case for which high frequency data like intraday data are available. In contrast, our proposed $\widehat{\theta}_{SI}$ can be applied to both high and low frequency data and is more applicable in practice. In addition, continuity of the sample path is also required for time-change which rules out jump-diffusion models and limits the applicability of the model while our estimator $\widehat{\theta}_{SI}$ only depends on the conditional moment restriction implied by the martingale property and works for models with jumps[9].

Moreover, the instantaneous conditional model in (5.37) studied in Park(2008) is only a univariate process since the time change technique does not apply for multivariate cases. However, our estimator $\widehat{\theta}_{SI}$ is able to estimate the multivariate version of the instantaneous conditional mean model $dX_t = b(X_t, \theta)dt + dU_t$, where $b(\cdot)$ is d-dimensional. In this case, the identification of the drift parameter θ_0 can be written as the following conditional moment restriction:

$$E\left[Z_t(\theta_0)|X_{\tau-1}\right] = 0 a.s.$$

for $\tau \geq 1$, where $Z_t(\theta)$ is a vector with components:

$$Z_t^i(\theta) = X_t^i - X_{t-\Delta}^i - \int_{t-\Delta}^{t} b_i(X_s;\theta)ds \tag{5.39}$$

for $i = 1, \cdots, d$. Similar to (5.36), the separate identification based estimator $\widehat{\theta}_{RV}$ for the multivariate instantaneous conditional mean model can be obtained by:

$$\begin{aligned}
&\widehat{\theta}_{RV} \\
= \quad & \arg\min_{\theta} \frac{1}{n-1}\sum_{l=2}^{n}\left|\frac{1}{n-1}\sum_{\tau=2}^{n} Z_\tau(\theta) 1\left(X_{\tau-1} \leq X_l\right)\right|^2 \\
= \quad & \arg\min_{\theta\in\Theta} \frac{1}{(n-1)^3}\sum_{i=1,\cdots,d}\left\{\frac{1}{n-1}\sum_{l=2}^{n}\left(\frac{1}{n-1}\sum_{\tau=2}^{n} Z_\tau^i(\theta) 1\left(X_{\tau-1} \leq X_l\right)\right)^2\right\} \qquad (5.40)
\end{aligned}$$

which covers the general multivariate semi-parametric diffusion model in (5.1).

[9]Hong, Lee and Song(2008) conduct an empirical study of interest rate process based on Park's(2008) instantaneous conditional mean model. To accommodate such features of the interest rate data as jumps and fixed sampling interval for which Park's(2008) time change based estimator does not apply, our separate identification based estimator is employed in their study.

5.4 Asymptotic Properties

5.4.1 Consistency

We shall prove the consistency of the two proposed estimators for the multivariate semi-parametric diffusion models in this section, i.e., the quadratic variation(covariation) based $\widehat{\theta}_{RV}$ in (5.32) and the separate identification based $\widehat{\theta}_{SI}$ in (5.40). Our plan is to first prove the consistency for the estimator $\widehat{\theta}_0$ in (5.22) by assuming the diffusion term is known and then deal with the unknown diffusion function estimated by the realized volatility(covariance) for $\widehat{\theta}_{RV}$. Moreover, the consistency of the second estimator $\widehat{\theta}_{SI}$ follows directly since it is just a special case of $\widehat{\theta}_0$. The following regularity conditions will be imposed throughout the whole section.

Assumption 5.4.1. $Z_t(\theta)$ is continuous on Θ; $|Z_t(\theta)| < Y$ with $E[Y] < \infty$; and $E[Z_t(\theta)|X_{t-\Delta}] = 0$ *a.s.* if and only $\theta = \theta_0$.

Assumption 5.4.2. $\begin{pmatrix} Z_t \\ X_{t-\Delta} \end{pmatrix}$ is ergodic and stationary

Assumption 5.4.3. $\Theta \subset \mathbb{R}^q$ is compact.

Assumptions 5.4.1–5.4.3 are standard in the GMM literature. Assumption 5.4.1 defines the model and identifies globally θ_0. It also establishes that the each component of $Z_t(\theta)$ is smooth in Θ. Notice that the assumptions concerning the existence of a bounding random variable Y and the compactness of Θ in Assumption 5.4.3 can be replaced by other assumptions imposing that for all $\theta \in \Theta$ there exists $\rho_\theta > 0$ such that $E\left[\sup_{\{\|\theta-\theta'\|\}\cap\Theta} |Z_t(\theta) - Z_t(\theta')|\right] < \infty$ and that $\liminf_{|\theta|\to\infty} E|Z_t(\theta) - Z_t(\theta')| > 0$. The second condition rules out redescending functions. Assumption 5.4.2 just restricts dependence and heterogeneity of the data. The consistency of $\widehat{\theta}_0$ is as follows:

Theorem 5.4.1: Under Assumptions 5.4.1-5.4.3, $\widehat{\theta}_0 \rightarrow^{a.s.} \theta_0$ as $n \rightarrow \infty$.

Then by dealing with the effect of replacing replacing $[X, X]$ by $\widehat{[X, X]}$ in θ_0 to obtain $\widehat{\theta}_{RV}$, the consistency of $\widehat{\theta}_{RV}$ is as follows:

Theorem 5.4.2: Under Assumptions 5.4.1-5.4.3, $\widehat{\theta}_{RV} \rightarrow^{a.s.} \theta_0$ as $n \rightarrow \infty$ and $h \rightarrow 0$.

The consistency of $\widehat{\theta}_{SI}$ can be delivered by a direct application of Theorem 5.4.1 for the case with separate identification condition.

Theorem 5.4.3: Suppose the corresponding assumptions for $Z_t(\theta_0, [X, X])$ to Assumptions 5.4.1-5.4.3 hold, then $\widehat{\theta}_{SI} \rightarrow^{p} \theta_0$ as $n \rightarrow \infty$.

Note that for the consistency of $\widehat{\theta}_{RV}$ in Theorem 5.4.2, both $n \rightarrow \infty$ and $h \rightarrow 0$ are required while for that of $\widehat{\theta}_{SI}$, only $n \rightarrow \infty$ has to be assumed. Therefore, the former requires high frequency data and the latter is applicable to both high and low frequency data.

5.4.2 Asymptotic Normality

The asymptotic normality of the proposed estimators will be proved in this section. Still the plan is to first prove that for the estimator $\widehat{\theta}_0$ in (5.22) by assuming the diffusion term is known and then show that the difference between $\widehat{\theta}_0$ and $\widehat{\theta}_{RV}$ is asymptotic negligible while for $\widehat{\theta}_{SI}$ as a special case of $\widehat{\theta}_0$, we directly apply the asymptotic normality of $\widehat{\theta}_0$.In order to obtain the asymptotic normality, the following additional assumptions are required.

Assumption 5.4.4: $Z_t(\theta)$ is first order continuously differentiable in a neighborhood of θ_0, Θ_0 and satisfies $E\left[\sup_{\theta\in\Theta_0}\left|\frac{\partial}{\partial\theta'}Z_t(\theta)\right|\right] < \infty$ where $\frac{\partial}{\partial\theta'}Z_t(\theta)$ is a $d' \times q$ matrix and $|\cdot|$ is the norm by reordering $\frac{\partial}{\partial\theta'}Z_t(\theta)$as a vector.

Assumption 5.4.5: $\theta_0 \in int\,(\Theta)$.

Assumption 5.4.6: $E\left[|Z_t(\theta_0)|^4\,||X_{t-\Delta}||^{1+\beta}\right] < \infty$ for some $\beta > 0$.

Assumption 5.4.7: The density of the conditioning variables given the past information is bounded and continuous.

Assumption 5.4.4 is a standard smoothness assumption. Assumption 5.4.5 is also standard in the literature. Assumptions 5.4.6 and 5.4.7 restrict the dependence of the conditioning variables with respect to the past. Notice that under independence, Assumption 5.4.7 can be relaxed to $E\,|Z_t(\theta_0)|^2 < \infty$ and Assumption 5.4.7 can be deleted, similarly to Stute (1997). Hence, for the independence case, no assumption concerning $X_{t-\Delta}$ would be required.

Let $\cdot H(x) \equiv E\left[\frac{\partial}{\partial\theta'}Z_t(\theta)1\,(X_{\tau-1} \leq x)\right]$, a $d' \times q$ matrix. The asymptotic normality of $\widehat{\theta_0}$ is stated as follows:

Theorem 5.4.4: Under Assumptions 5.4.1-5.4.7,

$$\sqrt{n}\left(\widehat{\theta_0} - \theta_0\right) \to^d N(0, \mathcal{V})$$

where

$$\mathcal{V} \;\equiv\; \mathcal{M}_q^{-1}\Omega_q\mathcal{M}_q^{-1}$$

$$\begin{aligned}\mathcal{M}_q &\equiv \int \cdot H'(x) \cdot H(x) dP_{X_2}(x) \\ \Omega_q &\equiv \int\int \cdot H'(x_1)\Gamma(x_1, x_2) \cdot H(x_2) dP_{X_2}(x_1) dP_{X_2}(x_2)\end{aligned}$$

are $q \times q$ matrices, $\Gamma(x_1, x_2)$ is a $d' \times d'$ matrix for each (x_1, x_2) in $\mathbb{R}^d \times \mathbb{R}^d$, with the (r, s)-th element

$$\Gamma_{r,s}(x_1, x_2) \equiv E\left[Z_t^r(\theta_0) Z_t^s(\theta_0) 1(X_{t-\Delta} \leq x_1 \wedge x_2)\right].$$

For the asymptotic normality of $\widehat{\theta}_{RV}$, $h \to 0$ is needed since the quadratic variation(covariation) is estimated by the realized volatility(covariance). Let $\cdot H(x, [X, X]) \equiv E\left[\frac{\partial}{\partial\theta'} Z_t(\theta, , [X, X]) 1(X_{t-\Delta} \leq x)\right]$, a $2 \times q$ matrix. The asymptotic normality of $\widehat{\theta}_{RV}$ is stated as follows:

Theorem 5.4.5: Suppose the corresponding assumptions for $Z_t(\theta, [X, X])$ to Assumptions 5.4.1-5.4.7 to $Z_t(\theta)$ hold and $h = n^{-1/2-\alpha}$ for some $\alpha > 0$,

$$\sqrt{n}\left(\widehat{\theta}_{RV} - \theta_{RV}\right) \to^d N(0, \mathcal{V}_{RV}), as n \to \infty and h \to 0$$

where

$$\begin{aligned}\mathcal{V}_{RV} &\equiv \mathcal{M}_{1q}^{-1} \Omega_{1q} \mathcal{M}_{1q}^{-1} \\ \mathcal{M}_{1q} &\equiv \int \cdot H'(x, [X, X]) \cdot H(x, [X, X]) dP_{X_2}(x) \\ \Omega_{1q} &\equiv \int\int \cdot H'(x_1, [X, X]) \Gamma_1(x_1, x_2) \cdot H(x_2, [X, X]) dP_{X_2}(x_1) dP_{X_2}(x_2)\end{aligned}$$

$q \times q$ matrices, $\Gamma_1(x_1, x_2)$ is a 2×2 matrix for each (x_1, x_2) in $\mathbb{R}^2$, with the (r, s)-th element

$$\Gamma_1^{r,s}(x_1, x_2) \equiv E\left[Z_t^r(\theta_{RV}, [X, X]) Z_t^s(\theta_{RV}, [X, X]) 1(X_{t-\Delta} \leq x_1 \wedge x_2)\right]$$

A simple consistent estimator of $\mathcal{V}_{RV}$ is its sample analog $\widehat{\mathcal{V}}_{RV} \equiv \widehat{\mathcal{M}}_{1q}^{-1}\widehat{\Omega}_{1q}\widehat{\mathcal{M}}_{1q}^{-1}$ with

$$
\begin{aligned}
\widehat{\mathcal{M}}_{1q} &\equiv \frac{1}{n-1}\sum_{l=2}^{n} \cdot H'_{n-1}(X_l, \widehat{[X,X]}) \cdot H'_{n-1}(X_l, \widehat{[X,X]}) \\
\widehat{\Omega}_{1q} &\equiv \frac{1}{(n-1)^2}\sum_{l=2}^{n}\sum_{l'=2}^{n} \cdot H'_{n-1}(X_l, \widehat{[X,X]}) \\
&\quad \Gamma_{1,n-1}\left(X_l, X_{l'}, \widehat{[X,X]}\right) \cdot H_{n-1}(X_{l'}, \widehat{[X,X]})
\end{aligned}
\tag{5.41}
$$

where

$$
\cdot H_{n-1}(x, \widehat{[X,X]}) \equiv \frac{1}{n-1}\sum_{\tau=2}^{n} \frac{\partial}{\partial\theta'} Z_\tau(\widehat{\theta}_{RV}, \widehat{[X,X]}) 1\left(X_{\tau-1} \leq x\right)
$$

and $\Gamma_{1,n-1}\left(X_l, X_{l'}, \widehat{[X,X]}\right)$ is a a 2×2 matrix for each (x_1, x_2) in $\mathbb{R}^2$, with the (r,s)-th element

$$
\begin{aligned}
&\Gamma_{1,n-1}^{r,s}\left(x_1, x_2, \widehat{[X,X]}\right) \\
\equiv\ & \frac{1}{n-1}\sum_{\tau=2}^{n} Z_\tau^r(\widehat{\theta}_{RV}, \widehat{[X,X]}) Z_\tau^s(\widehat{\theta}_{RV}, \widehat{[X,X]}) 1\left(X_{\tau-1} \leq x_1 \wedge x_2\right).
\end{aligned}
$$

For the asymptotic normality of $\widehat{\theta}_{SI}$, only $n \to \infty$ is needed and a direct application of Theorem 5.4.4 gives us:

Theorem 5.4.6: Under Assumptions 5.4.1-5.4.7,

$$
\sqrt{n}\left(\widehat{\theta}_{SI} - \theta_{SI}\right) \to^d N(0, \mathcal{V}_{SI})
$$

where

$$
\begin{aligned}
\mathcal{V}_{SI} &\equiv \mathcal{M}_{2q}^{-1}\Omega_{2q}\mathcal{M}_{2q}^{-1} \\
\mathcal{M}_{2q} &\equiv \int \cdot H'(x) \cdot H(x) dP_{X_2}(x) \\
\Omega_{2q} &\equiv \int\int \cdot H'(x_1)\Gamma\left(x_1, x_2\right) \cdot H(x_2) dP_{X_2}(x_1) dP_{X_2}(x_2)
\end{aligned}
$$

are $q \times q$ matrices, $\Gamma(x_1, x_2)$ is a $d' \times d'$ matrix for each (x_1, x_2) in $\mathbb{R}^d \times \mathbb{R}^d$, with the (r, s)-th element

$$\Gamma_{r,s}(x_1, x_2) \equiv E\left[Z_t^r(\theta_0) Z_t^s(\theta_0) 1\left(X_{t-\Delta} \leq x_1 \wedge x_2\right)\right]$$

Here in order to perform statistical inference, the matrix $\mathcal{V}_{SI}$ needs to be estimated consistently. A simple consistent estimator is its sample analog $\widehat{\mathcal{V}}_{SI} \equiv \widehat{\mathcal{M}}_{2q}^{-1} \widehat{\Omega}_{2q} \widehat{\mathcal{M}}_{2q}^{-1}$ with

$$\begin{aligned}
\widehat{\mathcal{M}}_{2q} &\equiv \frac{1}{n-1} \sum_{l=2}^{n} \cdot H'_{n-1}(X_l) \cdot H'_{n-1}(X_l) \\
\widehat{\Omega}_{2q} &\equiv \frac{1}{(n-1)^2} \sum_{l=2}^{n} \sum_{l'=2}^{n} \cdot H'_{n-1}(X_l) \Gamma_{n-1}\left(X_l, X_{l'}\right) \cdot H_{n-1}(X_{l'})
\end{aligned} \tag{5.42}$$

where

$$\cdot H_{n-1}(x) \equiv \frac{1}{n-1} \sum_{\tau=2}^{n} \frac{\partial}{\partial \theta'} Z_\tau(\widehat{\theta}_{SI}) 1\left(X_{\tau-1} \leq x\right)$$

and $\Gamma_{n-1}(X_l, X_{l'})$ is a a $d' \times d'$ matrix for each (x_1, x_2) in $\mathbb{R}^d \times \mathbb{R}^d$, with the (r, s)-th element

$$\Gamma_{n-1}^{r,s}(x_1, x_2) \equiv \frac{1}{n-1} \sum_{\tau=2}^{n} Z_\tau^r(\widehat{\theta}_{SI}) Z_\tau^s(\widehat{\theta}_{SI}) 1\left(X_{\tau-1} \leq x_1 \wedge x_2\right).$$

5.5 Simulation Studies

In this section, we shall conduct a comprehensive simulation study to check the finite sample performances of our proposed estimators. We first discuss about the simulation design.

5.5.1 Simulation Design

To have a complete understanding of how our proposed estimators perform for different model settings, four sets of models will be considered. The first set of models we shall examine are mainly diffusion models with a linear drift function, $dX_t = \kappa(\alpha - X_t)dt + \sigma(X_t)\,dW_t$ with $\theta = (\kappa, \alpha)$, for which both our proposed estimators $\widehat{\theta}_{RV}$ and $\widehat{\theta}_{SI}$ and the OLS estimator in Ait-Sahalia (1996b) are consistent. Three different specifications of the diffusion function $\sigma(\cdot)$ are considered to check whether the finite sample performances are sensitive to them:

- DGP 5.5.1: Vasicek (1977) Model

$$dX_t = \kappa(\alpha - X_t)dt + \sigma dW_t \tag{5.43}$$

with $\left(\kappa, \alpha, \sigma^2\right)$=(0.85837,0.089102,0.002185) and (0.214592,0.089102,0.000546) for the low and high persistent dependence cases respectively.

- DGP 5.5.2: CIR (Cox, Ingersoll and Ross, 1985) Model

$$dX_t = \kappa(\alpha - X_t)dt + \sigma\sqrt{X_t}dW_t \tag{5.44}$$

where $(\kappa, \alpha, \sigma^2)$ = (0.89218, 0.090495, 0.032742).

- DGP 5.5.3: CKLS (Chan, Karolyi, Longstaff and Sanders, 1992) Model:

$$dX_t = \kappa(\alpha - X_t)dt + \sigma X_t^{\rho} dW_t \tag{5.45}$$

where $(\kappa, \alpha, \sigma^2, \rho) = (0.0972, 0.0808, 0.52186, 1.46)$.

For this set of models, α is the long run mean and κ is the speed of mean reversion. The smaller κ is, the stronger the serial dependence in $\{X_t\}$ is, and consequently, the slower the process converges to the long run mean. We are particularly interested in the possible impact of dependent persistence in $\{X_t\}$ on the finite sample performances of the proposed estimators. Therefore, we follow Hong and Li (2005) and Pritsker(1998) to change κ and σ^2 in the same proportion so that the marginal density is unchanged; namely,

$$p(x, \theta) = \frac{1}{\sqrt{2\pi\sigma_s^2}} \exp\left[-\frac{(x-\alpha)^2}{2\sigma_s^2}\right]$$

where the stationary variance $\sigma_s^2 = \sigma^2/(2\kappa) = 0.01226$. In this way, we can focus on the impact of dependent persistence. Both low and high levels of dependent persistence are considered by adopting the same parameter values as those in Hong and Li (2005) and Pritsker (1998): $\left(\kappa, \alpha, \sigma^2\right)$ =(0.85837, 0.089102, 0.002185) and (0.214592, 0.089102, 0.000546) for the low and high persistent dependence cases respectively. The parameter values for the CIR model are also taken from Pritsker (1998) and those for the CKLS model from Ait-Sahalia's (1999) estimates of real interest rate data.

Both of our proposed estimators $\widehat{\theta}_{RV}$ and $\widehat{\theta}_{SI}$ will be compared to the OLS estimator in Ait-Sahalia (1996b) which is based on the following exact regression(see Ait-Sahalia (1996b) for the derivation):

$$E\left[X_{t+\Delta}|X_t\right] = \alpha + e^{-\kappa\Delta}(X_t - \alpha),\ t = \Delta, 2\Delta, \cdots, n\Delta \tag{5.46}$$

The OLS estimator $\widehat{\theta}_{OLS}$ obtained from (5.46)[10] is consistent under the linear drift specification[11]. The regression eqnarray (5.46) is also a property for the discrete sampled data which is derived by properties of the continuous time model(see Ait-Sahalia (1996a), page 532-533 for the derivation). It is not a result from discretizing the continuous time model and hence is free of the discretization bias. It is therefore applicable both to high frequency data and to low frequency data.

The second set of semi-parametric diffusion models are those with nonlinear drift specifications, for which both $\widehat{\theta}_{RV}$ and $\widehat{\theta}_{SI}$ are applicable while the OLS estimator in (6.4) does not work since the regression in (5.46) is not valid for the diffusion models with a nonlinear drift as discussed above. We consider two such models in the term structure of interest rates literature:

- DGP 5.5.4: Inverse-Feller Model(Ahn and Gao, 1999):

$$dX_t = X_t[\kappa - \beta X_t]dt + \sigma X_t^{3/2} dW_t \tag{5.47}$$

where $(\kappa, \alpha, \sigma^2)$ = (3.4387, 1.1361, 1.4209).

- DGP 5.5.5: Ait-Sahalia's (1996a) Nonlinear Drift Model:

[10] Ait-Sahalia (1996b) actually proposes a two step FGLS estimator by nonparametrically estimating the diffusion function in the second step. Here we only consider the exact OLS based on the regression eqnarray (5.46) for simplicity.

[11] As discussed in Aït-Sahalia (1996b) and Kristensen (2008b), the conditional mean $E\left[X_{(\tau+1)\Delta}|X_{\tau\Delta}\right]$ would generally involve the diffusion term when the drift term is nonlinear.

$$dX_t = (\kappa_{-1}X_t^{-1} + \kappa_0 + \kappa_1 X_t + \kappa_2 X_t^2)dt + \sigma X_t^{\rho} dW_t \tag{5.48}$$

where $(\kappa_{-1}, \kappa_0, \kappa_1, \kappa_2, \sigma^2, \rho)$=(0.0065, -0.3582, 6.6653, -35.4326, 3.2710, 5.2390).

The Inverse-Feller model in DGP4 proposed by Ahn and Gao(1999) is originally in the form of $dX_t = X_t[\kappa - \left(\sigma^2 - \kappa\alpha\right)X_t]dt + \sigma X_t^{3/2} dW_t$. We re-parameterize the model here for simplicity by setting $\beta = \sigma^2 - \kappa\alpha$. From the simplified form in (5.47), we can see that the drift is essentially a quadratic function of X_t without a constant term. It is the quadratic term $-\beta X_t^2$ that introduces the nonlinearity. For Ait-Sahalia's(1996a) general nonlinear drift model, the nonlinearity of the drift comes from two terms: $\kappa_2 X_t^2$ and $\kappa_{-1}X_t^{-1}$ which makes it more general. In fact, it can be observed by comparing (5.47) and (5.48) that Ahn and Gao's (1999) Inverse-Feller model is a special case of Ait-Sahalia's (1996a) nonlinear drift model. The former can be obtained by setting $\kappa_0 = \kappa_{-1} = 0$ and $\rho = 3/2$. The parameter values for the Inverse-Feller model are taken from Ahn and Gao (1999)[12] while those for Ait-Sahalia's (1996a) nonlinear drift model from Kristensen's (2008a) MLE estimates using daily observations of the Eurodollar interest rate data.

The third set of models are the stochastic volatility models and jump-diffusion models which do not belong to the semi-parametric diffusion class but are covered by the instantaneous conditional mean model in (5.37). Therefore, only $\widehat{\theta}_{SI}$ is applicable here and $\widehat{\theta}_{RV}$ is not consistent at all. These model are important in term-structure literature since it has been recognized that stochastic volatility helps improving the short-rate modeling while there has been strong

[12] Chen and Hong(2008) found some typos in the parameter values of Ahn and Gao's(1999) inverse-feller model by private correspondence and corrected therm. Here we choose the parameter values used by them.

evidence that jumps play an important role in capturing the dynamics of interest rates. For example, Andersen and Lund (1997) estimate the two factor stochastic volatility model and found that the stochastic volatility factor vastly improves goodness of fit and Johannes (2004) analyzes the role of jumps in continuous-time short interest rate models and the results show that jumps are both economically and statistically important. Two specific models are considered respectively here:

- DGP 5.5.6: Stochastic Volatility (SV) model in Andersen and Lund (1997) :

$$\begin{aligned} dX_t &= \kappa(\alpha - X_t)dt + \sigma_t X_t^{\gamma} dW_t, \\ d\log\sigma_t^2 &= \kappa_2(\alpha_2 - \log\sigma_t^2)dt + \xi dB_t, \end{aligned} \tag{5.49}$$

  where W_t and B_t are two independent standard Brownian Motions, $\kappa_1 = 0.1633$, $\alpha_1 = 0.0595$, $\gamma = 0.5443$, $\kappa_2 = 1.0397$, $\alpha_2 = -6.3599$, and $\xi = 1.2719$.

- DGP 5.5.7, Affine Jump-Diffusion model in Duffie, Pan and Singleton (2000):

$$dX_t = \kappa(\alpha - X_t)dt + \sigma\sqrt{X_t}dW_t + JdN_t, \tag{5.50}$$

where J is the random jump size which follows a $N\left(\mu_J, \sigma_J^2\right)$ distribution, N_t is a Poisson process with arrival intensity λ, W_t is a standard Brownian Motion, the diffusion and jump processes are independent of each other and are also independent of jump size J, $\kappa = 0.8542$, $\alpha = 0.0330$, $\sigma = 0.0173$, $\mu_J = 0.0004$, $\sigma_J = 0.0058$ and $\lambda = 5$.

The specification of SV model in DGP6 implies mean reversion for X_t as well as the stochastic (log-)volatility. The parameter we shall estimate is $\theta = (\kappa, \alpha)$

and the values are taken from Andersen and Lund's (1997) estimation results for weekly data. Comparing (5.49) with (5.45) tells us that the SV model in DGP6 is actually an extension of the CKLS model in DGP3 with a more general specification of the instantaneous volatility by letting σ be time-varying. For the Jump-Diffusion model in DGP 5.5.7, we assume that the coefficients are bounded and sufficient regularity conditions are satisfied so that a unique, strong solution to (5.49) exists (see Protter (2005) for details about the regularity conditions). The parameter values are taken from the estimation results of Das (2002) for daily interest rate data except that λ is set to 5 which corresponds to five jumps a year (Carrasco, Chernov, Florens and Ghysels, 2002).

The fourth set of models are multivariate diffusion models for which both $\widehat{\theta}_{RV}$ and $\widehat{\theta}_{SI}$ are consistent. In fact, according to our best knowledge, our proposed estimators may be the first consistent estimators for the multivariate semi-parametric diffusion models.

- DGP 5.5.8: Bivariate Ornstein-Ulenbeck(O-U) model

$$d\begin{bmatrix} X_{1t} \\ X_{2t} \end{bmatrix} = \begin{bmatrix} \kappa_{11} & 0 \\ \kappa_{21} & \kappa_{22} \end{bmatrix}\begin{bmatrix} X_{1t} \\ X_{2t} \end{bmatrix} dt + \begin{bmatrix} \sigma_{11} & 0 \\ 0 & \sigma_{22} \end{bmatrix} d\begin{bmatrix} W_{1t} \\ W_{2t} \end{bmatrix}$$

  with W_{1t} and W_{2t} two independent Brownian Motions and $\left(\kappa_{11}, \kappa_{21}, \kappa_{22}, \sigma_{11}^2, \sigma_{22}^2\right)$ =(-0.1117, 1.1138, -1.1637, 0.000546, 0.002185).

- DGP 5.5.9: Bivariate CIR model

$$\begin{aligned} d\begin{bmatrix} X_{1t} \\ X_{2t} \end{bmatrix} &= \left(\begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} + \begin{bmatrix} \kappa_{11} & \kappa_{12} \\ \kappa_{21} & \kappa_{22} \end{bmatrix}\begin{bmatrix} X_{1t} \\ X_{2t} \end{bmatrix}\right) dt \\ &\quad + \begin{pmatrix} \sigma_{11}\sqrt{X_{1t}} & 0 \\ 0 & \sigma_{22}\sqrt{X_{2t}} \end{pmatrix} d\begin{bmatrix} W_{1t} \\ W_{2t} \end{bmatrix} \end{aligned}$$

with W_{1t} and W_{2t} two independent Brownian Motions and (κ_{11}, κ_{12}, κ_{21}, κ_{22}, α_1, α_2, σ_{11}^2, σ_{22}^2)=(-0.7, 0.3, 1.2, -0.8, 0.56, 0.48, 0.002, 0.001)

Both of the two models above are two-factor affine diffusion term structure models(see Dai and Singleton (2000) and Ait-Sahalia and Kimmel (2009)). For the Bivariate O-U model, the transitional density does not admit an explicit form unless $\kappa_{21} = 0$ under which it follows Gaussian distribution(Duffee, 2002) while for the Bivariate CIR model to have a closed-form transition density, we need $\kappa_{21} = \kappa_{12} = 0$ under which it is equal to the product of two non-central chi-squared marginal distributions (Ait-Sahalia and Kimmel, 2009). The parameter values in the drift functions are the same as those in Chapter 2.

5.5.2 Simulation Results

We now present the results of our simulation studies. For each parameterization of the models considered, we simulate 1000(the number of replications) data sets of a random sample $\{X_t\}_{t=\Delta}^{n\Delta}$ at the monthly frequency ($\Delta = 1/12$) for n =250, 500, and 1000 respectively. These sample sizes correspond to around up to 80 years of monthly data. Since our quadratic variation(covariation) based estimator $\widehat{\theta}_{RV}$ requires the calculation of realized volatility(covariance) for each monthly interval, we actually generate data sets at the daily frequency with $m = 30$, $h = \Delta/m = 1/360$ and 5 observations each day. We discard the first 4 data points and save the fifth as the daily observation for each day and then discard the first 29 data points and save the thirtieth as the monthly observation for each month. The initial value is either generated by the known stationary distribution of X_t or set equal to the average interest rate level of the interest rate data set in Aït-Sahalia (1996b). Given a value X_t, we generate $X_{t+\Delta}$ either from the transition density when it is available in closed-form like that in Vasicek (1977) model or according to the so-called Euler scheme when it is not like the CKLS model. To eliminate the effect of initial values, we further generate 1000 more observations before starting the real simulation. The data generating schemes just described are actually employed for all the models we consider in this section and we shall not repeat them in the following.

For the first set of models with a linear drift function, $\widehat{\theta}_{RV}$, $\widehat{\theta}_{SI}$ and the OLS estimator $\widehat{\theta}_{OLS}$ in Ait-Sahalia(1996b) are reported in Table 5.1 and 5.2 with the average bias and mean squared error (MSE). We can observe from Table 5.1 for Vasicek model that both $\widehat{\kappa}_{RV}$ and $\widehat{\kappa}_{SI}$ have similar finite sample performances to those of $\widehat{\kappa}_{OLS}$ for the mean reverting parameter κ in terms of not only bias but also MSE, with $\widehat{\theta}_{RV}$ performing slightly better than the other two. This is actually true for both high and low persistent cases. Moreover all the three estimators tend to be rather imprecise when the sample size is small. For example, the relative bias of all the three estimators is around 50% even when the sample size n=500.

This is not very surprising since it has been well documented in the literature that estimation of the drift parameters can incur large bias and/or variability, see for instance Ball and Torous (1996) and Phillips and Yu (2005). It is the case for virtually all the commonly used estimation approaches including the maximum likelihood estimation for a full parametric diffusion model. Indeed, as reported in Phillips and Yu (2005) and Tang and Chen (2009), the maximum likelihood estimator for the mean reverting parameter can have more than 200% relative bias even the processes are observed monthly for more than 10 years.

Table 5.1: **Comparison of SI and RV-based estimators with OLS for the Vasicek Model**

Parameter Estimate	DGP 5.5.1 Vasicek Model					
	n=250		n=500		n=1000	
	Bias	MSE	Bias	MSE	Bias	MSE
	High Persistence: (κ, α) =(0.214592, 0.089102)					
$\widehat{\kappa}_{OLS}$	0.2267	0.1212	0.1121	0.0317	0.0558	0.0117
$\widehat{\alpha}_{OLS}$	0.0019	0.0012	0.0003	0.0003	0.0001	0.0001
$\widehat{\kappa}_{SI}$	0.2263	0.1338	0.1057	0.0363	0.0560	0.0151
$\widehat{\alpha}_{SI}$	0.0044	0.0055	0.0040	0.0037	0.0008	0.0001
$\widehat{\kappa}_{RV}$	0.1990	0.1176	0.0921	0.0329	0.0480	0.0126
$\widehat{\alpha}_{RV}$	0.0056	0.0033	0.0031	0.0003	0.0008	0.0001
	Low Persistence: (κ, α) =(0.85837, 0.089102)					
$\widehat{\kappa}_{OLS}$	0.2041	0.1771	0.1057	0.0682	0.0493	0.0258
$\widehat{\alpha}_{OLS}$	0.0002	0.0001	-0.0002	$7.35{\cdot}10^{-5}$	0.0002	$3.29{\cdot}10^{-5}$
$\widehat{\kappa}_{SI}$	0.2216	0.2594	0.0967	0.0884	0.0439	0.0385
$\widehat{\alpha}_{SI}$	0.0017	0.0002	0.0007	$9.11{\cdot}10^{-5}$	0.0004	$4.49{\cdot}10^{-5}$
$\widehat{\kappa}_{RV}$	0.2011	0.2121	0.1002	0.0896	0.0378	0.0338
$\widehat{\alpha}_{RV}$	0.0015	0.0001	0.0010	$7.82{\cdot}10^{-5}$	0.0007	$3.86{\cdot}10^{-5}$

Notes: (i), The model is DGP 5.5.1, Vasicek Model(1977), $dX_t = \kappa(\alpha - X_t)dt + \sigma dW_t$ with $\left(\kappa, \alpha, \sigma^2\right)$=(0.85837, 0.089102, 0.002185) and (0.214592, 0.089102 0.000546) for the low and high persistent dependence cases respectively. (ii), The number of replications is 1000.

Table 5.2: **Comparison of SI and RV-based estimators with OLS for the CIR and CKLS Models**

Parameter Estimate	DGP 5.5.2 CIR Model					
	n=250		n=500		n=1000	
	Bias	MSE	Bias	MSE	Bias	MSE
	True Parameters: (κ, α) = (0.89218, 0.090495)					
$\widehat{\kappa}_{OLS}$	0.2598	0.2302	0.1111	0.0770	0.0599	0.0369
$\widehat{\alpha}_{OLS}$	0.0002	0.0001	0.0017	$8.40{\cdot}10^{-5}$	$-2.08{\cdot}10^{-6}$	$4.64{\cdot}10^{-5}$
$\widehat{\kappa}_{SI}$	0.1303	0.1881	0.0713	0.0878	0.0347	0.0381
$\widehat{\alpha}_{SI}$	0.0046	0.0003	0.0016	0.00012	0.00093	$5.59{\cdot}10^{-5}$
$\widehat{\kappa}_{RV}$	0.1382	0.1803	0.0778	0.0877	0.0305	0.0386
$\widehat{\alpha}_{RV}$	0.0031	0.0002	0.0012	0.0001	0.0007	$5.34{\cdot}10^{-5}$
Parameter Estimate	DGP 5.5.3 CKLS Model					
	True Parameters: (κ, α) = (0.0972, 0.0808)					
$\widehat{\kappa}_{OLS}$	0.3472	0.1967	0.2325	0.0781	0.1560	0.0344
$\widehat{\alpha}_{OLS}$	0.0006	0.0045	0.0003	0.0020	-0.0008	0.0007
$\widehat{\kappa}_{SI}$	0.1777	0.1072	0.0639	0.0332	0.0252	0.0115
$\widehat{\alpha}_{SI}$	0.0330	1.2713	0.0012	0.1606	-0.0024	0.2037
$\widehat{\kappa}_{RV}$	0.1763	0.1036	0.0689	0.0319	0.0187	0.0124
$\widehat{\alpha}_{RV}$	-0.0083	0.0580	0.0031	0.2924	0.0006	0.1598

Notes: (i), The models are DGP 5.5.2, CIR model with a linear drift and $dX_t = \kappa(\alpha - X_t)dt + \sigma\sqrt{X_t}dW_t$ where $(\kappa, \alpha, \sigma^2)$ = (0.89218, 0.090495, 0.032742); DGP 5.5.3, CKLS model with a linear drift and $dX_t = \kappa(\alpha - X_t)dt + \sigma X_t^{\rho} dW_t$ where $(\kappa, \alpha, \sigma, \rho)$ = (0.0972, 0.0808, 0.722399, 1.46); (ii), The number of replications is 1000.

And we just find the similar problem for estimating semi-parametric diffusion models. The reason for such poor performances of the estimators is that the finite sample bias is effectively at the order of the time length T which is usually pretty small instead of the sample size n; see Tang and Chen (2009) for detailed discussions.

Another observation is that the estimators are truly affected by the persistence of the data generating processes, with better finite sample performances under low than under high persistent cases as can be seen by the fact that the relative bias of the mean reverting parameter estimator is around 100% in the high persistent case but is only 25% when the persistence is low for n=250. Lastly, the OLS estimator for the long-run mean parameter seems to be performing better than our $\widehat{\alpha}_{RV}$ and $\widehat{\alpha}_{SI}$. However, both $\widehat{\alpha}_{RV}$ and $\widehat{\alpha}_{SI}$ are actually already very precise and do not incur large bias, as evidenced by the 5% relative bias even when the sample size is as small as 250 for the high persistence case.

For the linear drift model with more complicated volatility dynamics than Vasicek model, i..e, the CIR and CKLS models, qualitatively similar results are observed from Table 5.2 with the sharp exception that for the mean-reverting parameter, both $\widehat{\kappa}_{RV}$ and $\widehat{\kappa}_{SI}$ are performing much better than the OLS estimator $\widehat{\kappa}_{OLS}$. The bias of the former is about half of the latter for the CIR model and is only 1/7 for the CKLS model when the sample size is 1000. In this sense, it seems our proposed estimators are more robust to the volatility dynamics.

For the second set of models with nonlinear drift specifications, only $\widehat{\theta}_{RV}$ and $\widehat{\theta}_{SI}$ are reported in Table 5.3 and the OLS estimator in (5.46) does not apply here. In general, the separate identification based estimators are performing better than the quadratic variation (covariation) based estimators, as is most ob-

Table 5.3: **SI and RV based estimators for the Inverse-Feller and Ait-Sahalia's Nonlinear Drift Models**

Parameter Estimate	DGP 5.5.4 Inverse-Feller Model Model					
	n=250		n=500		n=1000	
	Bias	MSE	Bias	MSE	Bias	MSE
	True Parameters: (κ, β) = (3.4387, 1.1361)					
$\widehat{\kappa}_{SI}$	0.1873	0.9643	0.1234	0.4818	0.1062	0.2498
$\widehat{\beta}_{SI}$	0.1103	0.3314	0.0634	0.1611	0.0525	0.0894
$\widehat{\kappa}_{RV}$	0.8242	8.9764	0.4087	2.8659	0.0973	2.0774
$\widehat{\beta}_{RV}$	0.5075	1.4000	0.3173	0.6487	0.1575	0.4073
Parameter	**DGP 5.5.5 Ait-Sahalia's Nonlinear Drift Model**					
Estimate	True Parameters: $(\kappa_{-1}, \kappa_0, \kappa_1, \kappa_2)$=(0.0065, -0.3582, 6.6653, -35.4326)					
$\widehat{\kappa}_{-1,SI}$	-0.0026	$9.21\cdot10^{-5}$	-0.0017	0.0003	-0.0016	0.0017
$\widehat{\kappa}_{0,SI}$	0.0711	0.0839	0.0453	0.2922	0.0431	1.5542
$\widehat{\kappa}_{1,SI}$	-0.6540	8.4883	-0.3937	29.5096	-0.3725	15.5492
$\widehat{\kappa}_{2,SI}$	1.9927	94.2680	1.1211	32.6895	1.0522	17.0601
$\widehat{\kappa}_{-1,RV}$	-0.0029	$8.81\cdot10^{-5}$	-0.0019	0.0003	-0.0016	0.0001
$\widehat{\kappa}_{0,RV}$	0.0830	0.0809	0.0527	0.3155	0.0424	0.1802
$\widehat{\kappa}_{1,RV}$	-0.7748	8.1705	-0.4680	31.8551	-0.3664	18.0293
$\widehat{\kappa}_{2,RV}$	2.3958	90.5405	1.3682	35.2751	1.0356	19.7771

Notes: (i), The models are DGP 5.5.4, Inverse-Feller model, with a nonlinear drift and $dX_t = \left(\kappa X_t - \beta X_t^2\right) dt + \sigma X_t^{3/2} dW_t$ where $(\kappa, \beta, \sigma^2)$ = (3.4387, 1.1361, 1.4209), DGP 5.5.5, Ait-Sahalia's Nonlinear Drift Model, with a general nonlinear drift and $dX_t = (\kappa_{-1} X_t^{-1} + \kappa_0 + \kappa_1 X_t + \kappa_2 X_t^2) dt + \sigma X_t^{\rho} dW_t$ where $(\kappa_{-1}, \kappa_0, \kappa_1, \kappa_2, \sigma^2, \rho)$=(0.0065, -0.3582, 6.6653, -35.4326, 3.2710, 5.2390). (ii), The number of replications is 1000.

vious for the Inverse-Feller model with the sample size 250 for which the bias of $\widehat{\kappa}_{RV}$ is almost 4 times bigger than that of $\widehat{\kappa}_{SI}$. However, the performance of $\widehat{\kappa}_{RV}$ is improving very quickly when the sample size is increasing. For example, $\widehat{\kappa}_{RV}$ has almost the same bias as $\widehat{\kappa}_{SI}$ when the sample size is reaching 1000. For Ait-Sahalia's general nonlinear drift model, $\widehat{\theta}_{RV}$ and $\widehat{\theta}_{SI}$ have pretty similar performances and no dominating results are observed for all sample sizes.

For the third set of models, i.e., SV and Jump-Diffusion models, only $\widehat{\theta}_{SI}$ is reported in Table 5.4 Table and $\widehat{\theta}_{RV}$ is not consistent since they do not belong to the semi-parametric diffusion class. For the mean-reverting parameter κ in the SV model, we observe from Table 5.4 that our separate identification based estimator is very imprecise when the sample size is 250 but is indeed improving fast with the increase of n. As discussed earlier, the big finite sample bias is actually due to the high persistence in the data generating process since the true value of κ is as small as 0.1633. The decreased finite sample bias from 100% to 20% when the sample size is increase from 250 to 1000 implies that our estimator is indeed robust to the complicated dynamics of stochastic volatility. Similarly, the small and decreased bias of our estimator $\widehat{\kappa}_{SI}$ with the increasing sample size for the Jump-Diffusion model show that the proposed estimator is also robust to the jumps and consistent under this more general setup.

For the fourth set of multivariate semi-parametric diffusion models, the results of both $\widehat{\theta}_{RV}$ and $\widehat{\theta}_{SI}$, which according to our best knowledge may be the first consistent estimators, are summarized in Table 5.5. For the Bivariate O-U model with constant volatility coefficients, the separate identification based estimators are performing better than the quadratic covariation based estimators for the mean reverting parameter κ_{11} of the first component process while worse

Table 5.4: **SI estimators for the Stochastic Volatility and Jump-Diffusion Models**

Parameter Estimate	**DGP 5.5.6 Stochastic Volatility Model**					
	n=250		n=500		n=1000	
	Bias	MSE	Bias	MSE	Bias	MSE
	True Parameters: (κ_1, α_1) = (0.1633, 0.0595)					
$\widehat{\kappa}_{1,SI}$	0.2148	0.1325	0.0850	0.0372	0.0399	0.0129
$\widehat{\alpha}_{1,SI}$	0.0065	0.0068	0.0047	0.0007	0.0016	$9.18{\cdot}10^{-5}$
Parameter Estimate	**DGP 5.5.7 Jump-Diffusion Model**					
	True Parameters: (κ, α) = (0.8542, 0.0330)					
$\widehat{\kappa}_{SI}$	0.1996	0.2169	0.1336	0.0990	0.0770	0.0418
$\widehat{\alpha}_{SI}$	0.0029	$2.47{\cdot}10^{-5}$	0.0025	$1.39{\cdot}10^{-5}$	0.0026	$9.83{\cdot}10^{-6}$

Notes: (i,) DGP 5.5.6 is the stochastic volatility model in Andersen and Lund (1997), with $\kappa_1 = 0.1633$, $\alpha_1 = 0.0595$, $\gamma = 0.5443$, $\kappa_2 = 1.0397$, $\alpha_2 = -6.3599$, and $\xi = 1.2719$. (ii), DGP 5.5.7 is the affine jump-diffusion model $dX_t = \kappa(\alpha - X_t)dt + \sigma\sqrt{X_t}dW_t + JdN_t$, with $\kappa = 0.8542$, $\alpha = 0.0330$, $\sigma = 0.0173$, $\mu_J = 0.0004$, $\sigma_J = 0.0058$ and $\lambda = 5$ (iii), The number of replications is 1000.

for both the mean reverting parameter κ_{22} of the second component process and the parameter κ_{21} controlling the correlation between the two component processes. For the Bivariate CIR ,model with more complicated volatility dynamics, $\widehat{\theta}_{RV}$ has better performances than $\widehat{\theta}_{SI}$ for both the mean reverting parameters and parameters controlling the correlation between the two component processes while worse performances for the long-run mean parameters. But overall, when the sample size is as big as 1000, both of them are fairly precise.

Table 5.5: **SI and RV estimators for Bivariate O-U and CIR models**

Parameter Estimate	DGP 5.5.8 Bivariate O-U Model					
	n = 250		n = 500		n = 1000	
	Bias	MSE	Bias	MSE	Bias	MSE
	True Parameters $(\kappa_{11}, \kappa_{21}, \kappa_{22})$ =(-0.1117,1.1138,-1.1637)					
$\widehat{\kappa}_{11,SI}$	0.0118	0.0668	0.0074	0.0250	-0.0006	0.0111
$\widehat{\kappa}_{21,SI}$	-0.2253	0.4169	-0.2380	0.2610	-0.2094	0.1798
$\widehat{\kappa}_{22,SI}$	0.2053	0.3306	0.2211	0.2216	0.1800	0.1442
$\widehat{\kappa}_{11,RV}$	-0.0042	0.0219	0.0044	0.0293	-0.0052	0.0119
$\widehat{\kappa}_{21,RV}$	-0.2644	0.6633	-0.1794	0.4476	-0.1169	0.2079
$\widehat{\kappa}_{22,RV}$	0.1903	0.3415	0.0703	0.1576	0.0637	0.0810
Parameter	Under DGP 5.5.9 Bivariate CIR Model					
Estimate	True Parameters $(\kappa_{11}, \kappa_{12}, \kappa_{21}, \kappa_{22}, \alpha_1, \alpha_2)$ =(-0.7, 0.3, 1.2, -0.8, 0.56, 0.48)					
$\widehat{\kappa}_{11,SI}$	0.5424	0.4194	0.3059	0.2184	0.0757	0.0822
$\widehat{\kappa}_{12,SI}$	-0.2842	0.1151	-0.1641	0.0614	-0.0448	0.0231
$\widehat{\kappa}_{21,SI}$	1.7784	4.3616	0.481419	0.5972	0.1318	0.1153
$\widehat{\kappa}_{22,SI}$	-0.9966	1.3212	-0.2955	0.1906	-0.1010	0.0377
$\widehat{\alpha}_{1,SI}$	-0.1584	0.0377	-0.0733	0.0183	0.0028	0.0099
$\widehat{\alpha}_{2,SI}$	-0.2679	0.1941	0.0554	0.0346	0.1160	0.0246
$\widehat{\kappa}_{11,RV}$	-0.0731	0.2334	-0.0044	0.0966	-0.0068	0.0458
$\widehat{\kappa}_{12,RV}$	-0.1368	0.1269	-0.0973	0.0514	-0.0459	0.0213
$\widehat{\kappa}_{21,RV}$	-0.0425	0.2046	-0.0383	0.0779	-0.039609	0.0383
$\widehat{\kappa}_{22,RV}$	-0.0706	0.1040	-0.0157	0.0401	0.0077	0.0177
$\widehat{\alpha}_{1,RV}$	0.8995	2.7379	0.5045	0.8421	0.2523	0.2709
$\widehat{\alpha}_{2,RV}$	0.4769	1.6211	0.1939	0.5577	0.0778	0.1752

Notes: (i), DGP 5.5.8 is the Bivariate O-U model and DGP 5.5.9 is the Bivariate CIR model (ii), The number of replications is 1000.

5.6 Conclusion

In this study, two GMM type estimators of drift parameters are proposed for both univariate and multivariate semi-parametric diffusion models with unrestricted volatility. The conditional moment restriction, through which the estimators are constructed, follows from a characterization of diffusion processes based on the infinitesimal operator, which is equivalent to transition density in terms of identifying the complete dynamics. The infinitesimal operator of the diffusion process enjoys the nice property of being a closed-form expression of drift and diffusion terms in an essentially separate manner, which makes the proposed estimators robust to the mis-specification of the diffusion function.

The first estimator is obtained by integrating out the diffusion function via the quadratic variation(covariation), which is estimated by the realized volatility(covariance) in a first step using high frequency data. The second estimator is constructed based on the separate identification condition and is actually applicable for a general instantaneous conditional mean model in continuous time, which covers the stochastic volatility and jump diffusion models as special cases. Our estimators for both univariate and multivariate models are unified in the same framework and particularly easy-to-implement. In fact, they may be the first consistent estimators for multivariate semi-parametric diffusion models, according to our best knowledge. Many empirical applications are possible by our proposed estimators; for example, we can extend Ait-Sahalia's(1996b) approach for pricing interest rate derivatives from the original semi-parametric diffusion model with a linear parametric drift and nonparametric diffusion to one with a nonlinear drift.

APPENDIX A

APPENDIX OF CHAPTER 2

A.1 Some Preliminary Lemmas

Lemma A.1.1: Let $g(y)$ be a real valued function such that $E\left|u\left(y_j\right)\right| < \infty$ and Assumptions 2.2.2 and 2.2.6 hold. Then $E\sum_{j=1}^{n}\left|u\left(y_j\right)\right| w_{ij} \leq CE\left|u\left(y_1\right)\right|$, where the constant C only depends on the kernel.

Proof of Lemma A.1.1: The results follows from a straightforward extension of Devroye and Wagner (1980, Lemma 2, p. 233) to the strong mixing sequence. Q.E.D.

Lemma A.1.2: Suppose the Assumptions 2.2.2, 2.2.3, 2.2.6, and conditions (2.45)-(2.47) and (2.49) are satisfied. Then

$$\max_{1\leq i\leq n} T_{i,n}\left\|\sum_{j=1}^{n} u\left(y_j;\theta\right) w_{ij}\right\| = o_p\left(\sqrt{\frac{\ln n}{nb_n^{dm+2\varsigma}}}\right) + O\left(\frac{b_n^2}{b_n^{\varsigma}}\right)$$

Proof of Lemma A.1.2: By the triangle inequality,

$$\max_{1\leq i\leq n} T_{i,n}\left\|\sum_{j=1}^{n} u\left(y_j;\theta\right) w_{ij}\right\| = (1)_A + (1)_B$$

where

$$\begin{aligned}
(1)_A &= \max_{1\leq i\leq n} T_{i,n}\left\|\sum_{j=1}^{n} u\left(y_j;\theta\right) w_{ij} - E\left\{\sum_{j=1}^{n} u\left(y_j;\theta\right) w_{ij}\right\}\right\| \\
(1)_B &= \max_{1\leq i\leq n} T_{i,n}\left\|E\left\{\sum_{j=1}^{n} u\left(y_j;\theta\right) w_{ij}\right\}\right\|
\end{aligned}$$

For any $\varepsilon > 0$ and $c_n \downarrow 0$,

$$P\left\{\max_{1\leq i\leq n} T_{i,n}\left\|\sum_{j=1}^{n} u\left(y_j;\theta\right) w_{ij} - E\left\{\sum_{j=1}^{n} u\left(y_j;\theta\right) w_{ij}\right\}\right\| > \varepsilon\sqrt{\frac{\ln n}{nb_n^{dm+2\varsigma}}}\right\}$$

$$\leq P\left\{\sup_{x_i\in\mathbb{R}^{dm}} T_{i,n}\left\|\sum_{j=1}^{n} u\left(y_j;\theta\right) w_{ij}-E\left\{\sum_{j=1}^{n} u\left(y_j;\theta\right) w_{ij}\right\}\right\|>\varepsilon\sqrt{\frac{\ln n}{nb_n^{dm+2\varsigma}}}\right\}$$
$$\leq P\left\{\sup_{x_i\in\mathbb{R}^{dm}}\left\|\frac{1}{nb_n^{dm}}\sum_{j=1}^{n} u\left(y_j;\theta\right)\mathcal{K}_{ij}-E\left\{\frac{1}{nb_n^{dm}}\sum_{j=1}^{n} u\left(y_j;\theta\right)\mathcal{K}_{ij}\right\}\right\|>\varepsilon\sqrt{\frac{\ln n}{nb_n^{dm+2\varsigma}}}b_n^{\varsigma}\right\}$$

where the definition of $T_{i,n}$ is used in the last inequality. The Assumptions 2.2.2, 2.2.3, 2.2.6, and conditions (2.45)-(2.47) and (2.49) imply the sufficient conditions in Hansen (2008, Theorem 4, p. 732), which provides us the following uniform convergence rate:

$$\sup_{x_i\in\mathbb{R}^{dm}}\left\|\frac{1}{nb_n^{dm}}\sum_{j=1}^{n} u\left(y_j;\theta\right)\mathcal{K}_{ij}-E\left\{\frac{1}{nb_n^{dm}}\sum_{j=1}^{n} u\left(y_j;\theta\right)\mathcal{K}_{ij}\right\}\right\|=o_p\left(\sqrt{\frac{\ln n}{nb_n^{dm}}}\right)$$

Hence,

$$P\left\{(1)_A>\varepsilon\sqrt{\frac{\ln n}{nb_n^{dm+2\varsigma}}}\right\}\to 0asn\to\infty.$$

which implies $(1)_A = o_p\left(\sqrt{\ln n/\left(nb_n^{dm+2\varsigma}\right)}\right)$. A standard argument similar to that for Hansen (2008, eqnarray (25), p. 733) shows that $(1)_B = O\left(b_n^2/b_n^{\varsigma}\right)$. Therefore, the desired results is proved by combining these two convergence rates. Q.E.D.

Lemma A.1.3: Suppose the Assumptions 2.2.2, 2.2.3, 2.2.6, and conditions (2.45)-(2.48) are satisfied. Then

$$\max_{1\leq i\leq n}\left|\widehat{h}\left(x_i\right)-h\left(x_i\right)\right|=O_p\left(\sqrt{\frac{\ln n}{nb_n^{dm}}}+b_n^2\right)$$

Proof of Lemma A.1.3: Observe that

$$\max_{1\leq i\leq n}\left|\widehat{h}\left(x_i\right)-h\left(x_i\right)\right|\leq\sup_{x_i\in\mathbb{R}^{dm}}\left|\widehat{h}\left(x_i\right)-h\left(x_i\right)\right|$$

The Assumptions 2.2.2, 2.2.3, 2.2.6, and conditions (2.45)-(2.48) imply the sufficient conditions in Hansen (2008, Theorem 6, p.733) which says

$$\sup_{x_i\in\mathbb{R}^{dm}}\left|\widehat{h}\left(x_i\right)-h\left(x_i\right)\right|=O_p\left(\sqrt{\frac{\ln n}{nb_n^{dm}}}+b_n^2\right).$$

This proves the conclusion. Q.E.D.

Lemma A.1.4: Suppose Assumptions 2.2.2, 2.2.3, 2.2.6 and 2.2.8 and conditions (2.45)-(2.47) and (2.50) hold. Then

$$\max_{1\le i\le n}\sup_{\theta\in\mathcal{B}_0} T_{i,n}\left\|\sum_{j=1}^{n}\frac{\partial u\left(y_j;\theta\right)}{\partial\theta}w_{ij}-D\left(x_i;\theta\right)\frac{h\left(x_i\right)}{\widehat{h}\left(x_i\right)}\right\| = O\left(\frac{b_n^2}{b_n^{\varsigma}}\right)+o_p\left(\sqrt{\frac{\ln n}{nb_n^{dm+2\varsigma}}}\right)$$

Proof of Lemma A.1.4: By the Triangle inequality and the definitions of $\widehat{h}\left(x_i\right)$ and $T_{i,n}$

$$\begin{aligned}&\max_{1\le i\le n}\sup_{\theta\in\mathcal{B}_0} T_{i,n}\left\|\sum_{j=1}^{n}\frac{\partial u\left(y_j;\theta\right)}{\partial\theta}w_{ij}-D\left(x_i;\theta\right)\frac{h\left(x_i\right)}{\widehat{h}\left(x_i\right)}\right\|\\ \le\ &\max_{1\le i\le n}\sup_{\theta\in\mathcal{B}_0}\frac{T_{i,n}}{b_n^r}\left\|\frac{1}{nb_n^{dm}}\sum_{j=1}^{n}\frac{\partial u\left(y_j;\theta\right)}{\partial\theta}\mathcal{K}_{ij}-D\left(x_i;\theta\right)h\left(x_i\right)\right\|\le\left[(2)_A+(2)_B\right]/b_n^{\varsigma}\end{aligned}$$

where

$$\begin{aligned}(2)_A &= \sup_{\theta\times x_i\in\mathcal{B}_0\times\mathbb{R}^{dm}}\left\|\frac{1}{nb_n^{dm}}\sum_{j=1}^{n}\frac{\partial u\left(y_j;\theta\right)}{\partial\theta}\mathcal{K}_{ij}-E\left\{\frac{1}{nb_n^{dm}}\sum_{j=1}^{n}\frac{\partial u\left(y_j;\theta\right)}{\partial\theta}w_{ij}\right\}\right\|,\\ (2)_B &= \sup_{\theta\times x_i\in\mathcal{B}_0\times\mathbb{R}^{dm}}\left\|E\left\{\frac{1}{nb_n^{dm}}\sum_{j=1}^{n}\frac{\partial u\left(y_j;\theta\right)}{\partial\theta}\mathcal{K}_{ij}-D\left(x_i;\theta\right)h\left(x_i\right)\right\}\right\|.\end{aligned}$$

Now the conditions in Assumptions 2.2.2 and 2.2.8 (iv) imply $(2)_B = O\left(b_n^2\right)$ by a standard argument similar to that for Hansen (2008, eqnarray (25), p. 733). Furthermore, the Assumptions 2.2.2, 2.2.3, 2.2.6, and 2.2.8 and conditions (2.45)-(2.47) and (2.50) imply the sufficient conditions in Hansen (2008, Theorem 4, p. 732), which tells us that $(2)_A = o_p\left(\sqrt{\ln n/\left(nb_n^{dm}\right)}\right)$. Combining these bounds, the conclusion is proved. Q.E.D.

Lemma A.1.5: Suppose Assumptions 2.2.2, 2.2.3, 2.2.6, and 2.2.8 and conditions (2.45)-(2.47) and (2.51) hold. Then

$$\max_{1\le i\le n}\sup_{\theta\in\mathcal{B}_0} T_{i,n}\left\|\widehat{V}\left(x_i;\theta\right)-V\left(x_i;\theta\right)\right\|$$

$$= o_p\left(\sqrt{\frac{\ln n}{nb_n^{dm+2\varsigma}}}\right) + O\left(\frac{b_n^2}{b_n^\varsigma}\right) + O_p\left(\sqrt{\frac{\ln n}{nb_n^{dm+2\varsigma}}} + \frac{b_n^2}{b_n^\varsigma}\right)$$

where

$$\widehat{V}(x_i;\theta) = \frac{1}{nb_n^{dm}}\sum_{j=1}^{n} u\left(y_j;\theta\right) u\left(y_j;\theta\right)' \mathcal{K}_{ij}$$

Proof of Lemma A.1.5: By the triangle inequality and the definitions of $\widehat{h}(x_i)$ and $T_{i,n}$,

$$\max_{1\le i\le n}\sup_{\theta\in\mathcal{B}_0} T_{i,n}\left\|\sum_{j=1}^{n} u\left(y_j;\theta\right) u\left(y_j;\theta\right)' w_{ij} - V(x_i;\theta)\right\| \le (3)_A + (3)_B$$

where

$$\begin{aligned}
(3)_A &= \max_{1\le i\le n}\sup_{\theta\in\mathcal{B}_0} \frac{T_{i,n}}{b_n^\varsigma}\left\|\frac{1}{nb_n^{dm}}\sum_{j=1}^{n} u\left(y_j;\theta\right) u\left(y_j;\theta\right)' \mathcal{K}_{ij} - V(x_i;\theta)\,h(x_i)\right\| \\
(3)_B &= \max_{1\le i\le n}\sup_{\theta\in\mathcal{B}_0} \frac{T_{i,n}}{b_n^\varsigma}\left\|V(x_i;\theta)\right\|\left\|h(x_i) - \widehat{h}(x_i)\right\|
\end{aligned}$$

Observe further that $(3)_A \le ((3)_{A1} + (3)_{A2})/b_n^\varsigma$, where

$$\begin{aligned}
(3)_{A1} &= c\left\|\frac{1}{nb_n^{dm}}\sum_{j=1}^{n} u\left(y_j;\theta\right) u\left(y_j;\theta\right)' \mathcal{K}_{ij} - E\left\{\frac{1}{nb_n^{dm}}\sum_{j=1}^{n} u\left(y_j;\theta\right) u\left(y_j;\theta\right)' \mathcal{K}_{ij}\right\}\right\| \\
(3)_{A2} &= \sup_{\theta\times x_i\in\mathcal{B}_0\times\mathbb{R}^{dm}}\left\|E\left\{\frac{1}{nb_n^{dm}}\sum_{j=1}^{n} u\left(y_j;\theta\right) u\left(y_j;\theta\right)' \mathcal{K}_{ij}\right\} - V(x_i;\theta)\,h(x_i)\right\|
\end{aligned}$$

The Assumptions 2.2.2, 2.2.3, 2.2.6, and 2.2.8 and conditions (2.45)-(2.47) and (2.51) imply the sufficient conditions in Hansen (2008, Theorem 4, p. 732), which shows that $(3)_{A1} = o_p\left(\sqrt{\ln n/(nb_n^{dm})}\right)$. Moreover, the conditions in Assumptions A.2.2.6 and A.2.2.8 (iii) imply $(3)_{A2} = O\left(b_n^2\right)$ by a standard argument similar to that for Hansen (2008, eqnarray (25), p. 733). Hence, it follows that

$$(3)_A = o_p\left(\sqrt{\frac{\ln n}{nb_n^{dm+2\varsigma}}}\right) + O\left(\frac{b_n^2}{b_n^\varsigma}\right) \tag{A.1}$$

Finally, since $\sup_{\theta\times x_i\in\mathcal{B}_0\times\mathbb{R}^{dm}}\|V(x_i;\theta)\| < \infty$ by Assumption 2.2.8 (v),

$$(3)_B = O_p\left(\sqrt{\frac{\ln n}{nb_n^{dm+2\varsigma}}} + \frac{b_n^2}{b_n^\varsigma}\right) \tag{A.2}$$

by Lemma A.1.3. The desired result follows by combining (A.1) and (A.2). Q.E.D.

Lemma A.1.6: Under the conditions of Lemma A.1.4,

$$\begin{aligned}
&\max_{1\le i\le n} T_{i,n}\left\|\left\{\sum_{j=1}^{n} u\left(y_j;\theta\right)u\left(y_j;\theta\right)' w_{ij}\right\}^{-1} - V^{-1}\left(x_i;\theta\right)\right\| \\
= \; & o_p\left(\sqrt{\frac{\ln n}{nb_n^{dm+2\varsigma}}}\right) + O\left(\frac{b_n^2}{b_n^{\varsigma}}\right) + O_p\left(\sqrt{\frac{\ln n}{nb_n^{dm+2\varsigma}}} + \frac{b_n^2}{b_n^{\varsigma}}\right)
\end{aligned}$$

Proof of Lemma A.1.6: It is straightforward by applying Lemma A.1.4 and **KTA** (Lemma D.1, p.1710). Q.E.D.

Lemma A.1.7: Suppose the conditions of Theorem 2.2.2 hold. For each i and $\theta \in \mathcal{B}_0$,

$$\begin{aligned}
&T_{i,n}\nabla_\theta\lambda_i\left(\theta\right)' \\
= \; & \widehat{T}_{i,n}V^{-1}\left(x_i;\theta\right)D\left(x_i;\theta\right) + \widehat{T}_{i,n}M_{1,i}\left(\theta\right)D\left(x_i;\theta\right) + \widehat{T}_{i,n}E\left[d\left(y_i|x_i\right)\right]M_{2,i}\left(\theta\right) \\
& + M_{3,i}\left(\theta\right)\sum_{j=1}^{n} d\left(y_j\right)w_{ij} + M_{4,i}\left(\theta\right)
\end{aligned}$$

where $\widehat{T}_{i,n} = T_{i,n}h\left(x_i\right)/\widehat{h}\left(x_i\right)$, $M_{1,i}\left(\theta\right)$ is a $q \times q$ matrix such that $\max_{1\le i\le n}\sup_{\theta\in\mathcal{B}_0}\left\|M_{1,i}\left(\theta\right)\right\| = o_p\left(1\right)$, and $M_{2,i}\left(\theta\right)$, $M_{3,i}\left(\theta\right)$, and $M_{4,i}\left(\theta\right)$ are $q\times p$ matrices such that $\max_{1\le i\le n}\sup_{\theta\in\mathcal{B}_0}\left\|M_{k,i}\left(\theta\right)\right\| = o_p\left(1\right)$ for $k = 2,3,4$.

Proof of Lemma A.1.7: Differentiating both sides of (2.33) with respect to θ yields:

$$\begin{aligned}
&\nabla_\theta\lambda_i'\left(\theta\right)\sum_{j=1}^{n}\frac{w_{ij}u\left(y_j;\theta\right)u\left(y_j;\theta\right)'}{\left[1+\lambda_i\left(\theta\right)'u\left(y_j;\theta\right)\right]^2} \\
= \; & \sum_{j=1}^{n}\frac{w_{ij}\nabla_\theta u\left(y_j;\theta\right)'}{1+\lambda_i\left(\theta\right)'u\left(y_j;\theta\right)} - \sum_{j=1}^{n}\frac{w_{ij}u\left(y_j;\theta\right)\lambda_i\left(\theta\right)'\nabla_\theta u\left(y_j;\theta\right)'}{\left[1+\lambda_i\left(\theta\right)'u\left(y_j;\theta\right)\right]^2} \qquad \text{(A.3)}
\end{aligned}$$

We shall simply the three terms in (A.3). First,

$$\begin{aligned}
&\max_{1\le i\le n}\sup_{\theta\in\mathcal{B}_0} T_{i,n}\left\|\sum_{j=1}^{n}\frac{w_{ij}u\left(y_j;\theta\right)u\left(y_j;\theta\right)'}{\left[1+\lambda_i(\theta)'u\left(y_j;\theta\right)\right]^2}-V(x_i;\theta)\right\| \\
\le\;& O(1)\max_{1\le i\le n}\sup_{\theta\in\mathcal{B}_0} T_{i,n}\left\|\widehat{V}(x_i;\theta)-V(x_i;\theta)\right\|+o(1)\max_{1\le i\le n}\sup_{\theta\in\mathcal{B}_0} T_{i,n}\|V(x_i;\theta)\|=o_p(1)
\end{aligned}$$

where Assumption A.2.2.9 is used in the inequality and Lemma A.1.5 and $\max_{1\le i\le n}\sup_{\theta\in\mathcal{B}_0}\|V(x_i;\theta)\|<\infty$ which follows from Assumption 2.2.8 (ii) are used in the equality. By Assumption 2.2.8 (ii), this implies

$$T_{i,n}\left\{\sum_{j=1}^{n}\frac{w_{ij}u\left(y_j;\theta\right)u\left(y_j;\theta\right)'}{\left[1+\lambda_i(\theta)'u\left(y_j;\theta\right)\right]^2}\right\}^{-1}=T_{i,n}V^{-1}(x_i;\theta)+R_{1,i}(\theta) \tag{A.4}$$

where $R_{1,i}(\theta)$ is a $q\times q$ matrix such that $\max_{1\le i\le n}\sup_{\theta\in\mathcal{B}_0}\left\|R_{1,i}(\theta)\right\|=o_p(1)$. Second,

$$\begin{aligned}
&\max_{1\le i\le n}\sup_{\theta\in\mathcal{B}_0} T_{i,n}\left\|\sum_{j=1}^{n}\frac{w_{ij}\nabla_\theta u\left(y_j;\theta\right)'}{\left[1+\lambda_i(\theta)'u\left(y_j;\theta\right)\right]}-D(x_i;\theta)\frac{h(x_i;\theta)}{\widehat{h}(x_i;\theta)}\right\| \\
\le\;& O(1)\max_{1\le i\le n}\sup_{\theta\in\mathcal{B}_0} T_{i,n}\left\|\sum_{j=1}^{n}w_{ij}\nabla_\theta u\left(y_j;\theta\right)'-D(x_i;\theta)\frac{h(x_i;\theta)}{\widehat{h}(x_i;\theta)}\right\| \\
&+o(1)\max_{1\le i\le n}\sup_{\theta\in\mathcal{B}_0}\widehat{T}_{i,n}E\left[d(y_i)\,|x_i\right] \\
=\;& o_p(1)
\end{aligned}$$

where Assumption 2.2.9 is used in the inequality and Lemma A.1.4 and $\max_{1\le i\le n}\sup_{\theta\in\mathcal{B}_0}\|D(x_i;\theta)\|<\infty$ which follows from Assumption 2.2.8 (iii) are used in the equality. This implies

$$T_{i,n}\sum_{j=1}^{n}\frac{w_{ij}\nabla_\theta u\left(y_j;\theta\right)'}{\left[1+\lambda_i(\theta)'u\left(y_j;\theta\right)\right]}=\widehat{T}_{i,n}D(x_i;\theta)+\widehat{T}_{i,n}E\left[d(y_i)\,|x_i\right]R_{2,i}(\theta)+R_{3,i}(\theta) \tag{A.5}$$

where $R_{2,i}(\theta)$ and $R_{3,i}(\theta)$ are $q\times p$ matrices such that $\max_{1\le i\le n}\sup_{\theta\in\mathcal{B}_0}\left(\left\|R_{2,i}(\theta)\right\|+\left\|R_{3,i}(\theta)\right\|\right)=o_p(1)$. Finally, Assumptions 2.2.8 (iii) and 2.2.9 imply that

$$\left\|\sum_{j=1}^{n}\frac{w_{ij}u\left(y_j;\theta\right)\lambda_i(\theta)'\nabla_\theta u\left(y_j;\theta\right)'}{\left[1+\lambda_i(\theta)'u\left(y_j;\theta\right)\right]^2}\right\|\le o(1)\sum_{j=1}^{n}d\left(y_j\right)w_{ij}$$

which further yields

$$\sum_{j=1}^{n} \frac{w_{ij} u\left(y_j;\theta\right) \lambda_i(\theta)' \nabla_\theta u\left(y_j;\theta\right)'}{\left[1+\lambda_i(\theta)' u\left(y_j;\theta\right)\right]^2} = R_{4,i}(\theta) \sum_{j=1}^{n} d\left(y_j\right) w_{ij} \tag{A.6}$$

where $R_{4,i}(\theta)$ is a $q \times p$ matrices such that $\max_{1\le i\le n} \sup_{\theta\in\mathcal{B}_0} \left\|R_{4,i}(\theta)\right\| = o_p(1)$. The desired result is then proved by combining (A.3)-(A.6). Q.E.D.

Lemma A.1.8: Under the conditions of Lemma A.1.3 and Assumption A.2.2.9

$$\frac{1}{n}\sum_{j=1}^{n}\left(T_{i,n}-1\right) = o_p(1)$$

and

$$\max_{1\le i\le n} \widehat{T}_{i,n} = O_p(1)$$

Proof of Lemma A.1.8: Fist observe that

$$\begin{aligned}
\frac{1}{n}\sum_{j=1}^{n}\left(T_{i,n}-1\right) &= \frac{1}{n}\sum_{j=1}^{n} 1\left[\widehat{h}(x_i) < b_n^{\varsigma}\right] \\
&\le \frac{1}{n}\sum_{j=1}^{n} 1\left[h(x_i) < 2b_n^{\varsigma}\right] + \max_{1\le i\le} 1\left[\left|\widehat{h}(x_i)-h(x_i)\right| > b_n^{\varsigma}\right]
\end{aligned}$$

By the law of large numbers for strongly mixing sequences and the fact that $E1\left[h(x_i) < 2b_n^{\varsigma}\right] \to 0$, the second last term is $o_p(1)$. Lemma A.1.3 and Assumption 2.2.9 shows that the last term is also $o_p(1)$. Hence the first part is proved. Then consider

$$\max_{1\le i\le n} \widehat{T}_{i,n} \le \max_{1\le i\le n} T_{i,n} \frac{\left|h(x_i)-\widehat{h}(x_i)\right|}{\widehat{h}(x_i)} + 1 = O_p(1)$$

where Lemma A.1.3 is used in the equality. This proves the second part. Q.E.D.

A.2 Auxiliary Propositions for the Main Results

We first introduce some notations. Let $g(x_i;\theta) = E[u(y;\theta)|x_i] / \left(1 + \left\|E[u(y;\theta)|x_i]\right\|\right)$ which will be used latter in the proof of Theorem 2.2.1 to replace $\lambda_i(\theta)$. Obviously, $\|g(x;\theta)\| \leq 1$ for all (x,θ). For a constant $\widetilde{c} \in (0,1)$, define $C_n = \left\{y : \sup_{\theta\in\Theta} \|u(y;\theta)\| \leq \widetilde{c}n^{1/r}\right\}$ and $u_n(y;\theta) = 1\{y \in C_n\} u(y;\theta)$. Finally, let $q_n(x,y;\theta) = -\log\left[1 + n^{-1/r} g'(x;\theta) u_n(y;\theta)\right]$ and define

$$Q_n(\theta) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} T_{i,n} w_{ij} q_n\left(x_i, y_j;\theta\right)$$

Proposition A.2.1. Suppose Assumptions 2.2.2, 2.2.3, 2.2.5 2.2.6, and 2.2.10 and conditions (2.45)-(2.47) and (2.49) hold. Then

$$\sup_{\theta\in\Theta}\left|Q_n(\theta) - \widetilde{Q}_n(\theta)\right| = o_p\left(n^{-1/r}\right)$$

where

$$\widetilde{Q}_n(\theta) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} -T_{i,n} g'(x_i;\theta) E[u(y_i;\theta)|x_i]$$

Proof of Proposition A.2.1: By the mean-value theorem, for some $t \in (0,1)$,

$$q_n(x,y;\theta) = -n^{-1/r} g'(x;\theta) u(y;\theta) + R_n(t) \tag{A.7}$$

where

$$R_n(t) = n^{-1/r} g'(x;\theta) u(y;\theta)(1 - 1\{y \in C_n\}) + \frac{n^{-2/r}\|cu_n(y;\theta)\|^2}{2\left(1 - tn^{-1/r} g'(x;\theta) u_n(y;\theta)\right)^2}$$

(A.7) and the fact that $\|g(x_i;\theta)\| \leq 1$ imply that

$$\begin{aligned} & n^{1/r}\sup_{\theta\in\Theta}\left|Q_n(\theta) - \widetilde{Q}_n(\theta)\right| \\ \leq\ & \sup_{\theta\in\Theta}\frac{1}{n}\sum_{i=1}^{n} T_{i,n}\left\|\sum_{j=1}^{n} w_{ij} u\left(y_j;\theta\right) - E[u(y_i;\theta)|x_i]\right\| \\ & + n^{1/r}\sup_{\theta\in\Theta}\left\|\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} T_{i,n} w_{ij} R_n(t)\right\| \end{aligned} \tag{A.8}$$

By repeated applications of the Cauchy-Schwartz inequality,

$$|R_n(t)| \leq n^{-1/r} \sup_{\theta \in \Theta} |u(y;\theta)| (1 - 1\{y \in C_n\}) + \frac{1}{2(1-\widetilde{c})^2} n^{-2/r} \sup_{\theta \in \Theta} \|u(y;\theta)\|^2$$

which, combined with Lemma A.1.1 and $\max_{1\leq j\leq n} 1\left\{y_j \notin C_n\right\} = o_p(1)$, implies that

$$n^{1/r} \sup_{\theta \in \Theta} \left\| \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} T_{i,n} w_{ij} R_n(t) \right\| = o_p(1) \tag{A.9}$$

By the triangle inequality and the definitions of $\widehat{h}(x_i)$ and $T_{i,n}$,

$$\sup_{\theta \in \Theta} \frac{1}{n} \sum_{i=1}^{n} T_{i,n} \left\| \sum_{j=1}^{n} w_{ij} u\left(y_j;\theta\right) - E\left[u(y_i;\theta)|x_i\right] \right\| \leq ((4)_A + (4)_B + (4)_C)/b_n^{\varsigma} \tag{A.10}$$

where

$$\begin{aligned}
(4)_A &= \sup_{\theta \times x_i \in \Theta \times \mathbb{R}^{dm}} \left\| \frac{1}{nb_n^{dm}} \sum_{j=1}^{n} \mathcal{K}_{ij} u\left(y_j;\theta\right) - E\left\{ \frac{1}{nb_n^{dm}} \sum_{j=1}^{n} \mathcal{K}_{ij} u\left(y_j;\theta\right) \right\} \right\| \\
(4)_B &= \sup_{\theta \in \Theta} \left\| E\left\{ \frac{1}{nb_n^{dm}} \sum_{j=1}^{n} \mathcal{K}_{ij} u\left(y_j;\theta\right) \right\} - h(x_i) E\left[u(y_i;\theta)|x_i\right] \right\| \\
(4)_C &= \max_{1\leq i\leq n} \left| h(x_i) - \widehat{h}(x_i) \right| \frac{1}{n} \sum_{i=1}^{n} E\left\{ \sup_{\theta \in \Theta} \|u(y_i;\theta)\| \mid x_i \right\}
\end{aligned}$$

Since $(1/n)\sum_{i=1}^{n} E\{\sup_{\theta\in\Theta} \|u(y_i;\theta)\| \mid x_i\} = O_p(1)$ by Assumption 2.2.5, we have

$$(4)_C = O_p\left(\sqrt{\frac{\ln n}{nb_n^{dm}}} + b_n^2 \right) \tag{A.11}$$

by Lemma A.1.3. The conditions in Assumptions 2.2.6 and 2.2.8 (iii) imply

$$(4)_B = O\left(b_n^2\right) \tag{A.12}$$

by a standard argument similar to that for Hansen (2008, eqnarray (25), p. 733). Furthermore, Assumptions 2.2.2, 2.2.3, 2.2.6, and conditions (2.45)-(2.47) and (2.49) imply the sufficient conditions in Hansen (2008, Theorem 4, p. 732), which provides us the following uniform convergence rate:

$$(4)_A = o_p\left(\sqrt{\frac{\ln n}{nb_n^{dm}}} \right) \tag{A.13}$$

Combining (A.8)-(A.13), we have

$$
\begin{aligned}
& n^{1/r} \sup_{\theta\in\Theta} \left| Q_n(\theta) - \widetilde{Q}_n(\theta) \right| \\
= \;& o_p\left(\sqrt{\frac{\ln n}{n b_n^{dm+2\varsigma}}} \right) + O\left(b_n^2 / b_n^{\varsigma} \right) + O_p\left(\sqrt{\frac{\ln n}{n b_n^{dm+2\varsigma}}} + b_n^2 / b_n^{\varsigma} \right) + o_p(1) \\
= \;& o_p(1)
\end{aligned}
$$

where Assumption 2.2.10 is used in the second equality. This finishes the proof of Proposition A.2.1. Q.E.D.

Proposition A.2.2: Under the conditions of Lemmas A.1.2, A.1.5, and A.1.6,

$$
T_{i,n}\lambda_i(\theta_0) = T_{i,n}\widehat{V}(x_i;\theta_0)^{-1} \sum_{j=1}^{n} u\left(y_j;\theta_0\right) w_{ij} + T_{i,n} r_i
$$

where

$$
\max_{1\le i\le n} T_{i,n} \|r_i\| = o_p\left(\frac{n^{1/r} \ln n}{n b_n^{dm+2\varsigma}} \right) + O\left(\frac{n^{1/r} b_n^4}{b_n^{2\varsigma}} \right)
$$

Proof of Proposition A.2.2: By (2.33),

$$
0 = \sum_{j=1}^{n} u\left(y_j;\theta_0\right) w_{ij} - \widehat{V}(x_i;\theta_0)\,\lambda_i(\theta_0) + \sum_{j=1}^{n} \frac{w_{ij} u\left(y_j;\theta_0\right) \left[\lambda_i(\theta_0)' u\left(y_j;\theta_0\right)\right]^2}{1+\lambda_i(\theta_0)' u\left(y_j;\theta_0\right)}
$$

By Lemma A.1.5 and Assumption 2.2.8 (ii), $T_{i,n}\widehat{V}(x_i;\theta_0)$ is invertible w.p.a.1. Hence,

$$
T_{i,n}\lambda_i(\theta_0) = T_{i,n}\widehat{V}(x_i;\theta_0)^{-1} \sum_{j=1}^{n} u\left(y_j;\theta_0\right) w_{ij} + T_{i,n} r_i \tag{A.14}
$$

where

$$
r_i = \widehat{V}(x_i;\theta_0)^{-1} \sum_{j=1}^{n} \frac{w_{ij} u\left(y_j;\theta_0\right) \left[\lambda_i(\theta_0)' u\left(y_j;\theta_0\right)\right]^2}{1+\lambda_i(\theta_0)' u\left(y_j;\theta_0\right)}
$$

From (2.33), we also obtain

$$
T_{i,n} \sum_{j=1}^{n} \frac{w_{ij} \left[\lambda_i(\theta_0)' u\left(y_j;\theta_0\right)\right]^2}{1+\lambda_i(\theta_0)' u\left(y_j;\theta_0\right)} = T_{i,n} \sum_{j=1}^{n} w_{ij}\lambda_i(\theta_0)' u\left(y_j;\theta_0\right) \tag{A.15}
$$

which implies

$$\begin{aligned}
& T_{i,n} \|r_i\| / \left\| \widehat{V}(x_i;\theta_0)^{-1} \right\| \\
\leq\ & \max_{1\leq j\leq n} \left\| u\left(y_j;\theta_o\right)\right\| T_{i,n} \sum_{j=1}^{n} w_{ij}\lambda_i(\theta_0)' u\left(y_j;\theta_0\right) \\
=\ & o\left(n^{1/r}\right) T_{i,n} \sum_{j=1}^{n} w_{ij}\lambda_i(\theta_0)' u\left(y_j;\theta_0\right) \\
\leq\ & o\left(n^{1/r}\right) T_{i,n} \|\lambda_i(\theta_o)\| \left\{ o_p\left(\sqrt{\frac{\ln n}{nb_n^{dm+2\varsigma}}}\right) + O\left(\frac{b_n^2}{b_n^{\varsigma}}\right)\right\} \qquad \text{(A.16)}
\end{aligned}$$

where **KTA** (Lemma D.2, p. 1711) is used in the second equality and Lemma A.1.2 in the second inequality. Now let $\lambda_i(\theta_0) = \rho_i\xi_i$ where $\rho_i \geq 0$ and $\xi_i \in S^q$. Because

$$0 \leq 1 + \lambda_i(\theta_0)' u\left(y_j;\theta_0\right) \leq 1 + \rho_i(\theta_0)\left\|u\left(y_j;\theta_0\right)\right\| = 1 + \rho_i(\theta_0)\, o\left(n^{1/r}\right)$$

where **KTA** (Lemma D.2, p. 1711) is used again in the last step. Hence, (A.15) becomes

$$\frac{T_{i,n}\rho_i}{1+\rho_i o\left(n^{1/r}\right)} \leq \frac{T_{i,n}\sum_{j=1}^{n} w_{ij}\xi_i' u\left(y_j;\theta_0\right)}{\xi_i'\widehat{V}(x_i;\theta_o)\,\xi_i}$$

which further implies

$$\max_{1\leq i\leq n} T_{i,n}\rho_i = o_p\left(\sqrt{\frac{\ln n}{nb_n^{dm+2\varsigma}}}\right) + O\left(\frac{b_n^2}{b_n^{\varsigma}}\right) \qquad \text{(A.17)}$$

by Lemma A.1.2. Hence by (A.16) and (A.17)

$$\begin{aligned}
\max_{1\leq i\leq n} T_{i,n}\|r_i\| &= \max_{1\leq i\leq n}\left\|\widehat{V}(x_i;\theta_0)^{-1}\right\| \times \max_{1\leq i\leq n} T_{i,n}\left\|r_{1,i}\right\| / \left\|\widehat{V}(x_i;\theta_0)^{-1}\right\| \\
&= o_p\left(\frac{n^{1/r}\ln n}{nb_n^{dm+2\varsigma}}\right) + O\left(\frac{n^{1/r}b_n^4}{b_n^{2\varsigma}}\right)
\end{aligned}$$

where $\max_{1\leq i\leq n}\left\|\widehat{V}(x_i;\theta_0)^{-1}\right\| = O_p(1)$ which follows from Lemma A.1.6 is used in the second equality. This finishes the proof. Q.E.D.

Proposition A.2.3: Under the conditions of Theorem 2.2.2,

$$\sup_{\theta\in\mathcal{B}_0}\left\|-\frac{1}{n}\nabla_{\theta\theta}\mathbf{LELL}(\theta)-I(\theta)=o_p(1)\right\|$$

Proof of Proposition A.2.3: By (2.33) and (2.34),

$$-\nabla_{\theta}\mathbf{LELL}(\theta)=T_1(\theta)+T_2(\theta)+T_3(\theta)$$

where

$$\begin{aligned}
T_1(\theta) &= -\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{T_{i,n}w_{ij}\left[\nabla_{\theta}\left\{\lambda_i(\theta)'u\left(y_j;\theta\right)\right\}\right]\lambda_i(\theta_0)'\nabla_{\theta}u'\left(y_j;\theta\right)}{\left[1+\lambda_i(\theta)'u\left(y_j;\theta\right)\right]^2}\\
T_2(\theta) &= \sum_{i=1}^{n}\sum_{j=1}^{n}\frac{T_{i,n}w_{ij}\left[\nabla_{\theta}\lambda_i(\theta)\right]\nabla_{\theta}u'\left(y_j;\theta\right)}{1+\lambda_i(\theta)'u\left(y_j;\theta\right)}\\
T_3(\theta) &= \sum_{i=1}^{n}\sum_{j=1}^{n}\frac{T_{i,n}w_{ij}}{1+\lambda_i(\theta)'u\left(y_j;\theta\right)}\sum_{k=1}^{q}\left[\nabla_{\theta\theta}u^{(k)}\left(y_j;\theta\right)\right]\lambda_i^{(k)}(\theta)
\end{aligned}$$

The desired result follows from Propositions A.2.4-A.2.6. Q.E.D.

Proposition A.2.4: Under the conditions of Theorem 2.2.2, $\sup_{\theta\in\mathcal{B}_0}\|T_2(\theta)/n-I(\theta)\|=o_p(1)$

Proof of Proposition A.2.4: By the triangle inequality and definitions of $T_{i,n}$ and $\widehat{h}(x_i)$

$$\|T_2(\theta)/n-I(\theta)\|\leq(5)_A+(5)_B+(5)_C+(5)_D+(5)_E \tag{A.18}$$

where

$$\begin{aligned}
(5)_A &= \left\|\frac{T_2(\theta)}{n}-\frac{1}{n}\sum_{j=1}^{n}\widehat{T}_{i,n}\left[\nabla_{\theta}\lambda_i(\theta)\right]D(x_i;\theta)\right\|\\
(5)_B &= \left\|\frac{1}{n}\sum_{i=1}^{n}\widehat{T}_{i,n}\left[\nabla_{\theta}\lambda_i(\theta)\right]D(x_i;\theta)-\frac{1}{n}\sum_{i=1}^{n}\widehat{T}_{i,n}^2D'(x_i;\theta)V^{-1}(x_i;\theta)D(x_i;\theta)\right\|
\end{aligned}$$

$$(5)_C = \left\| \frac{1}{n} \sum_{i=1}^{n} T_{i,n} \frac{h^2(x_i) - \widehat{h}^2(x_i)}{\widehat{h}^2(x_i)} D'(x_i;\theta) V^{-1}(x_i;\theta) D(x_i;\theta) \right\|$$

$$(5)_D = \left\| \frac{1}{n} \sum_{i=1}^{n} (1 - T_{i,n}) D'(x_i;\theta) V^{-1}(x_i;\theta) D(x_i;\theta) \right\|$$

$$(5)_E = \left\| \frac{1}{n} \sum_{i=1}^{n} D'(x_i;\theta) V^{-1}(x_i;\theta) D(x_i;\theta) - I(\theta) \right\|$$

First, by (A.5),

$$(5)_A \leq \left\| \frac{1}{n} \sum_{i=1}^{n} \widehat{T}_{i,n} [\nabla_\theta \lambda_i(\theta)] E[d(y_i)|x_i] R_{2,i}(\theta) \right\| + \left\| \frac{1}{n} \sum_{i=1}^{n} T_{i,n} [\nabla_\theta \lambda_i(\theta)] R_{3,i}(\theta) \right\| \quad (A.19)$$

where $\max_{1\leq i\leq n} \sup_{\theta\in\mathcal{B}_0} \left(\left\|R_{2,i}(\theta)\right\| + \left\|R_{3,i}(\theta)\right\| \right) = o_p(1)$. Then Lemma A.1.8, $\sup_{(x_i,\theta)\in\mathbb{R}^{dm}\times\mathcal{B}_0} \left\|V(x_i;\theta)^{-1}\right\| < \infty$ which follows from Assumption A.2.2.8 (ii), and Lemma A.1.7 imply

$$\sup_{\theta\in\mathcal{B}_0} \frac{1}{n} \sum_{i=1}^{n} T_{i,n} \|[\nabla_\theta \lambda_i(\theta)]\|^2 = O_p(1) \, and \sup_{\theta\in\mathcal{B}_0} \frac{1}{n} \sum_{i=1}^{n} T_{i,n} \|[\nabla_\theta \lambda_i(\theta)]\| = O_p(1) \quad (A.20)$$

Then (A.19), (A.20), and applications of Cauchy-Schwartz inequality yield

$$(5)_A = o_p(1) \quad (A.21)$$

Applying the Cauchy-Schwartz and Jensen inequalities to Lemma A.1.7 again,

$$(5)_B = o_p(1) \quad (A.22)$$

Observe that

$$(5)_C \leq \frac{C}{b_n^{2\varsigma}} \max_{1\leq i\leq n} \left|\widehat{h}(x_i) - h(x_i)\right| \max_{1\leq i\leq n} \left|\widehat{h}(x_i) + h(x_i)\right| \frac{1}{n} \sum_{i=1}^{n} \sup_{\theta\in\mathcal{B}_0} \|D(x_i,\theta)\|^2 = o_p(1) \quad (A.23)$$

where Lemma A.1.3, Assumption 2.2.6, and $E\left\{\frac{1}{n}\sum_{i=1}^{n} \sup_{\theta\in\mathcal{B}_0} \|D(x_i,\theta)\|^2\right\} < \infty$ which follows from Assumption 2.2.8 (iii) are used in the equality. By Cauchy-Schwartz inequality,

$$(5)_D \leq \sqrt{\frac{1}{n} \sum_{i=1}^{n} (1 - T_{i,n}) \frac{C}{n} \sum_{i=1}^{n} \sup_{\theta\in\mathcal{B}_0} \|D(x_i,\theta)\|^4} = o_p(1) \quad (A.24)$$

where Lemma A.1.8, Assumption 2.2.7, and $E\left\{\frac{1}{n}\sum_{i=1}^{n}\sup_{\theta\in\mathcal{B}_0}\|D(x_i,\theta)\|^4\right\} < \infty$ which follows from Assumption 2.2.8 (iii) are used in the equality. Finally,

$$(5)_E = o_p(1) \tag{A.25}$$

follows from a uniform weak law of large numbers for strongly mixing sequences. The desired results follows from (A.18) and (A.21)-(A.25). Q.E.D.

Proposition A.2.5: Under the conditions of Theorem 2.2.2, $\sup_{\theta\in\mathcal{B}_0}\|T_1(\theta)/n\| = o_p(1)$

Proof of Proposition A.2.5: Observe that

$$T_1(\theta)/n = (6)_A + (6)_B \tag{A.26}$$

where

$$\begin{aligned}
(6)_A &= -\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{T_{i,n}w_{ij}}{\left[1+\lambda_i(\theta)' u\left(y_j;\theta\right)\right]^2}\left[\nabla_\theta u\left(y_j;\theta\right)\right]\lambda_i(\theta)\lambda_i(\theta)'\nabla_\theta u'\left(y_j;\theta\right)\\
(6)_B &= -\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{T_{i,n}w_{ij}}{\left[1+\lambda_i(\theta)' u\left(y_j;\theta\right)\right]^2}\left[\nabla_\theta\lambda_i(\theta)\right]u\left(y_j;\theta\right)\lambda_i(\theta)'\left[\nabla_\theta u\left(y_j;\theta\right)'\right]
\end{aligned}$$

Assumptions 2.2.8 (iii) and 2.2.9 and Lemma A.1.1 imply that

$$\sup_{\theta\in\mathcal{B}_0}\|(6)_A\| \le o(1)\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}w_{ij}d^2\left(y_j\right) = o_p(1) \tag{A.27}$$

Similarly,

$$\begin{aligned}
\sup_{\theta\in\mathcal{B}_0}\|(6)_B\| &\le o(1)\sup_{\theta\in\mathcal{B}_0}\frac{1}{n}\sum_{i=1}^{n}T_{i,n}\|\nabla_\theta\lambda_i(\theta)\|\sum_{j=1}^{n}w_{ij}d\left(y_j\right)\\
&\le o(1)\left\{\sup_{\theta\in\mathcal{B}_0}\frac{1}{n}\sum_{i=1}^{n}T_{i,n}\|\nabla_\theta\lambda_i(\theta)\|^2\right\}^{1/2}\left\{\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}w_{ij}d^2\left(y_j\right)\right\}^{1/2}\\
&= o_p(1)
\end{aligned} \tag{A.28}$$

where Cauchy-Schwartz and Jensen inequalities are used in the second inequality and (A.20) and Lemma A.1.1 are used in the equality. The desired result follows from (A.26)-(A.28). Q.E.D.

Proposition A.2.6: Under the conditions of Theorem 2.2.2, $\sup_{\theta\in\mathcal{B}_0}\|T_3(\theta)/n\| = o_p(1)$

Proof of Proposition A.2.6: Assumptions 2.2.8 (iii), 2.2.9, and 2.2.10, and Lemma A.1.1 imply that

$$\sup_{\theta\in\mathcal{B}_0}\|T_3(\theta)/n\| \leq o(1)\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij} l\left(y_j\right) = o_p(1)$$

Hence, the conclusion is proved. Q.E.D.

Proposition A.2.7: Under the conditions of Theorem 2, $n^{-1/2}A \to^d N(0, I(\theta_0))$, where

$$A = \sum_{i=1}^{n} T_{i,n}\left(\sum_{j=1}^{n} w_{ij}\frac{\partial u'\left(y_j;\theta_0\right)}{\partial\theta}\right)\widehat{V}^{-1}(x_i;\theta_0)\left(\sum_{j=1}^{n} w_{ij} u\left(y_j;\theta_0\right)\right)$$

Proof of Proposition A.2.7: Since A is a $p \times 1$ vector, we shall use Carmer-Wold device to prove the asymptotic normality. Let $\xi \in S^p$ be arbitrary and choose a small positive constant ϑ such that

$$\max_{1\leq i\leq n}\left|\widehat{h}(x_i) - h(x_i)\right|/\alpha_n = o_p(1), where \alpha_n \triangleq b_n^{\varsigma+\vartheta} \tag{A.29}$$

which is feasible in light of Lemma A.1.3. Define $T_{i,n}^{*} = 1\{h(x_i) \geq b_n^{\varsigma} - \alpha_n\}$ and then

$$T_{i,n}\left(1 - T_{i,n}^{*}\right) = 1\left\{\widehat{h}(x_i) \geq b_n^{\varsigma} and h(x_i) < b_n^{\varsigma} - \alpha_n\right\} \leq 1\left\{\max_{1\leq i\leq n}\left|\widehat{h}(x_i) - h(x_i)\right| > \alpha_n\right\} \to 0$$

by (A.29). This implies that

$$T_{i,n}\left(1 - T_{i,n}^{*}\right) = 0 for all 1 \leq i \leq n w.p.a.1 \tag{A.30}$$

Decompose

$$\xi' A = (7)_A + (7)_B + (7)_C + (7)_D + (7)_E + (7)_F \tag{A.31}$$

where

$$
\begin{aligned}
(7)_A &= \frac{1}{nb_n^{dm}} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{T_{i,n}^{*}}{E\left[\widehat{h}(x_i)\,|x_i\right]} \xi' J_1(x_i) P_1^{-1}(x_i) u\left(y_j;\theta_0\right) \mathcal{K}_{ij} \\
(7)_B &= \frac{1}{nb_n^{dm}} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{(T_{i,n}-1) T_{i,n}^{*}}{E\left[\widehat{h}(x_i)\,|x_i\right]} \xi' J_1(x_i) P_1^{-1}(x_i) u\left(y_j;\theta_0\right) \mathcal{K}_{ij} \\
(7)_C &= \frac{1}{nb_n^{dm}} \sum_{i=1}^{n} \sum_{j=1}^{n} T_{i,n} T_{i,n}^{*} \left\{ \frac{1}{\widehat{h}(x_i)} - \frac{1}{E\left[\widehat{h}(x_i)\,|x_i\right]} \right\} \xi' J_1(x_i) P_1^{-1}(x_i) u\left(y_j;\theta_0\right) \mathcal{K}_{ij} \\
(7)_D &= -\frac{1}{nb_n^{dm}} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{T_{i,n} T_{i,n}^{*}}{\widehat{h}(x_i)} \xi' J_1(x_i) P_1^{-1}(x_i) \\
&\quad \left\{ I_{q\times q} + \widehat{P}_2(x_i) P_1^{-1}(x_i) \right\}^{-1} \widehat{P}_2(x_i) P_1^{-1}(x_i) u\left(y_j;\theta_0\right) \mathcal{K}_{ij} \\
(7)_E &= \frac{1}{nb_n^{dm}} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{T_{i,n}\left(1-T_{i,n}^{*}\right)}{\widehat{h}(x_i)} \xi' J_1(x_i) \left\{ P_1(x_i) + \widehat{P}_2(x_i) \right\}^{-1} u\left(y_j;\theta_0\right) \mathcal{K}_{ij} \\
(7)_F &= \frac{1}{nb_n^{dm}} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{T_{i,n}}{\widehat{h}^2(x_i)} \xi' \widehat{J}_2(x_i) \widehat{V}(x_i;\theta_0)^{-1} u\left(y_j;\theta_0\right) \mathcal{K}_{ij}
\end{aligned}
$$

and

$$
\begin{aligned}
J_1(x_i) &= E\left\{ \frac{1}{nb_n^{dm}} \sum_{i=1}^{n} \mathcal{K}_{ij} \frac{\partial u\left(y_j;\theta_0\right)}{\partial \theta} \right| \left. x_i \right\} \\
\widehat{J}_2(x_i) &= \frac{1}{nb_n^{dm}} \sum_{i=1}^{n} \mathcal{K}_{ij} \frac{\partial u\left(y_j;\theta_0\right)}{\partial \theta} - E\left\{ \frac{1}{nb_n^{dm}} \sum_{i=1}^{n} \mathcal{K}_{ij} \frac{\partial u\left(y_j;\theta_0\right)}{\partial \theta} \right| \left. x_i \right\} \\
P_1(x_i) &= E\left\{ \frac{1}{nb_n^{dm}} \sum_{i=1}^{n} \mathcal{K}_{ij} u\left(y_j;\theta_0\right) u\left(y_j;\theta_0\right)' \right| \left. x_i \right\} \\
\widehat{P}_2(x_i) &= \frac{1}{nb_n^{dm}} \sum_{i=1}^{n} \mathcal{K}_{ij} u\left(y_j;\theta_0\right) u\left(y_j;\theta_0\right)' - E\left\{ \frac{1}{nb_n^{dm}} \sum_{i=1}^{n} \mathcal{K}_{ij} u\left(y_j;\theta_0\right) u\left(y_j;\theta_0\right)' \right| \left. x_i \right\}
\end{aligned}
$$

Now we shall analyze the terms in (A.37). First, observe that

$$
\begin{aligned}
&\left| n^{-1/2} (7)_F \right| \\
\le\ & \frac{n^{1/2}}{b_n^{2\varsigma}} \max_{1\le i\le n} \left\| \widehat{J}_2(x_i) \right\| \max_{1\le i\le n} T_{i,n} \left\| \widehat{V}^{-1}(x_i;\theta_0) \right\| \max_{1\le i\le n} \left\| \frac{1}{nb_n^{dm}} \sum_{i=1}^{n} \mathcal{K}_{ij} u\left(y_j;\theta_0\right) \right\| \\
\le\ & \frac{n^{1/2}}{b_n^{2\varsigma}} o_p\left(\sqrt{\frac{\ln n}{nb_n^{dm}}} \right) o_p\left(\sqrt{\frac{\ln n}{nb_n^{dm}}} \right) = o_p(1) \qquad \text{(A.32)}
\end{aligned}
$$

where $\max_{1\leq i\leq n}\left\|\widehat{J_2}(x_i)\right\| = o_p\left(\sqrt{\ln n/(nb_n^{dm})}\right)$ which can be shown as in Lemma A.1.4, boundedness of $\max_{1\leq i\leq n} T_{i,n}\left\|\widehat{V}^{-1}(x_i;\theta_0)\right\|$ which follows from Lemma A.1.5 and Assumption 2.2.7, and $\max_{1\leq i\leq n}\left\|\sum_{i=1}^{n}\mathcal{K}_{ij}u\left(y_j;\theta_0\right)/\left(nb_n^{dm}\right)\right\| = o_p\left(\sqrt{\ln n/(nb_n^{dm})}\right)$ which can be shown as in Lemma A.1.2 are used in the second inequality and Assumption 2.2.8 is used in the equality. Second, it follows from (A.31) that

$$(7)_E = 0 w.p.a.1 \tag{A.33}$$

Third, similar to (A.32), it can be shown that

$$\begin{aligned}
& \left|n^{-1/2}(7)_D\right| \\
\leq\ & \frac{n^{1/2}}{b_n^{2\varsigma}}\left\{\max_{1\leq i\leq n} T^*_{i,n}\left\|P_1^{-1}(x_i)h(x_i)\right\|\right\}^2 \max_{1\leq i\leq n} T^*_{i,n}\left\|\left\{I_{q\times q}+\frac{\widehat{P}_2(x_i)}{h(x_i)}P_1^{-1}(x_i)h(x_i)\right\}^{-1}\right\| \\
& \max_{1\leq i\leq n} T^*_{i,n}\left\|\frac{\widehat{P}_2(x_i)}{h(x_i)}\right\| \max_{1\leq i\leq n}\left\|\frac{1}{nb_n^{dm}}\sum_{i=1}^{n}\mathcal{K}_{ij}u\left(y_j;\theta_0\right)\right\|\frac{1}{n}\sum_{i=1}^{n}\left\|J_1(x_i)\right\| \\
\leq\ & \frac{n^{1/2}}{b_n^{2\varsigma}}\left\{O_p\left(\sqrt{\frac{\ln n}{nb_n^{dm+2\varsigma}}}+\frac{b_n^2}{b_n^{\varsigma}}\right)\right\}^2 o_p\left(\sqrt{\frac{\ln n}{nb_n^{dm}}}\right)o_p\left(\sqrt{\frac{\ln n}{nb_n^{dm}}}\right) \\
=\ & o_p(1)
\end{aligned} \tag{A.34}$$

and also

$$\begin{aligned}
& \left|n^{-1/2}(7)_C\right| \\
\leq\ & \frac{n^{1/2}}{b_n^{3\varsigma}}\max_{1\leq i\leq n}\left\|\widehat{h}(x_i)-E\left[\widehat{h}(x_i)|x_i\right]\right\|\max_{1\leq i\leq n} T^*_{i,n}\left\|P_1^{-1}(x_i)h(x_i)\right\| \\
& \max_{1\leq i\leq n}\left\|\frac{1}{nb_n^{dm}}\sum_{i=1}^{n}\mathcal{K}_{ij}u\left(y_j;\theta_0\right)\right\|\frac{1}{n}\sum_{i=1}^{n}\left\|J_1(x_i)\right\| \\
\leq\ & \frac{n^{1/2}}{b_n^{3\varsigma}}O\left(b_n^2\right)o_p\left(\sqrt{\frac{\ln n}{nb_n^{dm}}}\right)o_p\left(\sqrt{\frac{\ln n}{nb_n^{dm}}}\right) = o_p(1)
\end{aligned} \tag{A.35}$$

Fourth, an application of (A.30) again implies that

$$(7)_B = 0 w.p.a.1 \tag{A.36}$$

Last, observe that $(7)_A$ can be regarded as a degenerate U-statistic. Then by applying a similar argument as **KTA** (Lemma B.2, p.1696-1698) and employing the central limit theorem of U-statistics for absolute regular processes (Fan and Li, 1999), it can be shown that

$$n^{-1/2}(7)_A \to^d N\left(0, I\left(\theta_0\right)\right) \tag{A.37}$$

The desired result follows from (A.31)-(A.37). Q.E.D.

A.3 Proofs of the Main Results

Proof of Theorem 2.2.1: First note that $\widehat{\theta}_{LEL}$ maximizes the objective function

$$G_n(\theta) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} -T_{i,n}w_{ij}\log\left\{1 + \lambda_i'(\theta)\, u\left(y_j;\theta\right)\right\}$$

Then recall the definitions of $Q_n(\theta)$ and $\widetilde{Q}_n(\theta)$ at the beginning of Appendix A.2 and in Proposition A.2.1 respectively. It is easy to see that

$$G_n(\theta) \le Q_n(\theta) \tag{A.38}$$

for all θ by the optimality of $\lambda_i' s$ which follows from (2.36). Define

$$\overline{Q}_n(\theta) = \frac{1}{n^{1+1/r}}\sum_{i=1}^{n} -g\left(x_i;\theta\right)' E\left[u\left(y_i;\theta\right)|x_i\right]$$

which is the same as $\widetilde{Q}_n(\theta)$ except for the absence of the trimming factor $T_{i,n}$. Then by Cauchy-Schwartz inequality,

$$\begin{aligned}
& \sup_{\theta\in\Theta} n^{1/m}\left|\widetilde{Q}_n(\theta) - \overline{Q}_n(\theta)\right| \\
\le\ & \frac{1}{n}\sum_{i=1}^{n}\left(T_{i,n}-1\right)\sup_{\theta\in\Theta}\left\{\left\|g\left(x_i;\theta\right)\right\|\left\|E\left[u\left(y_i;\theta\right)|x_i\right]\right\|\right\} \\
\le\ & \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(T_{i,n}-1\right)}\sqrt{\frac{1}{n}\sum_{i=1}^{n}E\left[\sup_{\theta\in\Theta}\left\|g\left(x_i;\theta\right)\right\|^2 \mid x_i\right]} \\
=\ & o_p(1)
\end{aligned} \tag{A.39}$$

where Lemma A.1.8 and Assumption 2.2.5 are used in the equality. By (A.38), (A.39) and Proposition A.2.1,

$$\sup_{\theta\in\Theta} n^{1/m} G_n(\theta) \leq \sup_{\theta\in\Theta} n^{1/m} Q_n(\theta) = \sup_{\theta\in\Theta} n^{1/m} \overline{Q}_n(\theta) + o_p(1) \tag{A.40}$$

Next, we can apply a uniform law of large numbers to $n^{1/m}\overline{Q}_n(\theta)$ for strongly mixing sequences since the follow sufficient conditions are satisfied: first, $E[u(y_i;\theta)|x_i]$ is continuous in θ w.p.1 by Assumptions 2.2.5 and 2.2.7 (iii) and the Bounded Convergence Theorem. This implies

$$-g(x_i;\theta)' E[u(y_i;\theta)|x_i] = \frac{-\left\|E[u(y_i;\theta)|x_i]\right\|^2}{1+\left\|E[u(y_i;\theta)|x_i]\right\|}$$

is also continuous in θ w.p.1. Second, $E\left[\sup_{\theta\in\Theta}\left|-g(x_i;\theta)' E[u(y_i;\theta)|x_i]\right|\right]$ is finite under Assumption 2.2.7. Third, Θ is compact. Hence, the uniform law implies

$$\sup_{\theta\in\Theta}\left|n^{1/m}\overline{Q}_n(\theta) - E\{-g(x_i;\theta)' E[u(y_i;\theta)|x_i]\}\right| = o_p(1) \tag{A.41}$$

where $-E\{g(x_i;\theta)' E[u(y_i;\theta)|x_i]\}$ is continuous in θ and

$$\begin{aligned} & -E\{g(x_i;\theta)' E[u(y_i;\theta)|x_i]\} \\ \leq\ & -E\left\{\frac{1[x_i\in\mathcal{X}_\theta]\left\|E[u(y_i;\theta)|x_i]\right\|^2}{1+\left\|E[u(y_i;\theta)|x_i]\right\|}\right\} < 0 \end{aligned}$$

for $\theta \neq \theta_0$ by Assumption 2.2.2. Hence by the continuity of $-E\{g(x_i;\theta)' E[u(y_i;\theta)|x_i]\}$ and compactness of Θ, for each $\delta > 0$, there exists a strictly positive number $H(\delta)$ such that

$$\sup_{\theta\in\Theta\backslash\mathcal{B}(\theta_0,\delta)} n^{1/m} G_n(\theta) E\{-g(x_i;\theta)' E[u(y_i;\theta)|x_i]\} \leq -H(\delta) \tag{A.42}$$

Hence, by (A.40)-(A.42),

$$P\left\{\sup_{\theta\in\Theta\backslash\mathcal{B}(\theta_0,\delta)} n^{1/m} G_n(\theta) > -H(\delta)\right\} < \delta/2 \tag{A.43}$$

for sufficiently large n. Next, we analyze $G_n(\theta_\theta)$. It follows from Lemma A.1.2 and (A.17) that

$$
\begin{aligned}
G_n(\theta_\theta) &= -\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} T_{i,n} w_{ij} \log\left\{1+\lambda_i'(\theta_\theta)\, u\left(y_j;\theta_\theta\right)\right\} \\
&\geq -\frac{1}{n}\sum_{i=1}^{n} T_{i,n}\lambda_i'(\theta_\theta)\sum_{j=1}^{n} w_{ij} u\left(y_j;\theta_\theta\right) \\
&= \left\{o_p\left(\sqrt{\frac{\ln n}{n b_n^{dm+2\varsigma}}}\right)+O\left(\frac{b_n^2}{b_n^{\varsigma}}\right)\right\}\left\{o_p\left(\sqrt{\frac{\ln n}{n b_n^{dm+2\varsigma}}}\right)+O\left(\frac{b_n^2}{b_n^{\varsigma}}\right)\right\} \\
&\triangleq o_p\left(d_n^2\right)
\end{aligned}
$$

where Assumption 2.2.10 is used in the last step. Therefore,

$$
P\left\{G_n(\theta_\theta)/d_n^2 < H(\delta)\right\} < \delta/2 forsufficientlylargen \tag{A.44}
$$

Assumption 2.2.10 implies $d_n^2 n^{1/r} \downarrow 0$. Hence (A.43) and (A.44) together finish the proof. Q.E.D.

Proof of Theorem 2.2.2: By a Taylor expansion of $\nabla_\theta \mathbf{LELL}\left(\widehat{\theta}_{LEL}\right)=0$ which is the first order condition for (2.35),

$$
0 = n^{-1/2}\nabla_\theta \mathbf{LELL}(\theta_0) + \frac{1}{n}\nabla_{\theta\theta}\mathbf{LELL}(\theta^*)\, n^{1/2}\left(\widehat{\theta}_{LEL}-\theta_0\right) \tag{A.45}
$$

where θ^* is between $\widehat{\theta}_{LEL}$ and θ_0. Moreover, (2.33) implies

$$
-\nabla_\theta \mathbf{LELL}(\theta) = \sum_{i=1}^{n}\sum_{j=1}^{n}\frac{T_{i,n} w_{ij}\left[\nabla_\theta u'\left(y_j;\theta\right)\right]\lambda_i(\theta)}{1+\lambda_i(\theta)'\, u\left(y_j;\theta\right)} \tag{A.46}
$$

Then by (A.45), (A.46) and Proposition A.2.2,

$$
-n^{-1/2}\nabla_\theta \mathbf{LELL}(\theta_0) = n^{-1/2}(8)_A + n^{-1/2}(8)_B + n^{-1/2}(8)_C \tag{A.47}
$$

where

$$
(8)_A = A
$$

$$
\begin{aligned}
(8)_B &= \sum_{i=1}^{n} T_{i,n} \left(\sum_{j=1}^{n} w_{ij} \left\{ \frac{\partial u'\left(y_j;\theta_0\right)/\partial\theta}{1+\lambda_i\left(\theta_0\right)' u\left(y_j;\theta_0\right)} - \frac{\partial u'\left(y_j;\theta_0\right)}{\partial\theta} \right\} \right) \\
&\quad \widehat{V}^{-1}\left(x_i;\theta_0\right) \left(\sum_{j=1}^{n} w_{ij} u\left(y_j;\theta_0\right) \right) \\
(8)_C &= \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{T_{i,n} w_{ij} \nabla_\theta u\left(y_j;\theta_0\right) r_i}{1+\lambda_i\left(\theta_0\right)' u\left(y_j;\theta_0\right)}
\end{aligned}
$$

where A is defined in Proposition A.2.7. Since $\max_{1\le j\le n}\sup_{\theta\in\Theta}\left\|u\left(y_j;\theta\right)\right\| = o\left(n^{1/r}\right)$ w.p.1 **KTA** (Lemma D.2, p.1711) and Assumption A.2.2.3, Assumption 2.2.7 implies

$$
\max_{1\le i,j\le n}\sup_{\theta\in\Theta}\left\|\lambda_i' u\left(y_j;\theta\right)\right\| = o\left(1\right)
$$

which further implies

$$
\max_{1\le i,j\le n}\sup_{\theta\in\Theta}\frac{1}{\left|1+\lambda_i' u\left(y_j;\theta\right)\right|} = O\left(1\right) w.p.1 \tag{A.48}
$$

Hence,

$$
\begin{aligned}
\left\|n^{-1/2}(8)_C\right\| &\le n^{1/2} O\left(1\right) \max_{1\le i\le n} T_{i,n}\left\|r_i\right\| \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} d\left(y_j\right) w_{ij} \\
&\le n^{1/2}\left\{ o_p\left(\frac{n^{1/r}\ln n}{n b_n^{dm+2\varsigma}}\right) + O\left(\frac{n^{1/r} b_n^4}{b_n^{2\varsigma}}\right)\right\} = o_p\left(1\right)
\end{aligned} \tag{A.49}
$$

where (A.48) and Assumption 2.2.8 (iii) are used in the first inequality, Proposition A.2.2 and Lemma A.1.1 are used in the second inequality, and Assumption 2.2.10 is used in the last step. Furthermore,

$$
\begin{aligned}
&\left\|n^{-1/2}(8)_B\right\| \\
&= O_p\left(1\right)\max_{1\le i\le n} T_{i,n}\left\|\sum_{j=1}^{n} w_{ij} u\left(y_j;\theta_0\right)\right\| \\
&\quad \frac{1}{\sqrt{n}}\sum_{i=1}^{n} T_{i,n}\sum_{j=1}^{n} w_{ij}\left\| \frac{\partial u'\left(y_j;\theta_0\right)/\partial\theta}{1+\lambda_i\left(\theta_0\right)' u\left(y_j;\theta_0\right)} - \frac{\partial u'\left(y_j;\theta_0\right)}{\partial\theta}\right\|
\end{aligned}
$$

$$
\begin{aligned}
&= O_p(1)\max_{1\le i\le n} T_{i,n}\left\|\sum_{j=1}^{n} w_{ij}u\left(y_j;\theta_0\right)\right\| \\
&\quad \frac{1}{\sqrt{n}}\sum_{i=1}^{n} T_{i,n}\left\|\lambda_i(\theta_0)\right\|\sum_{j=1}^{n} w_{ij}d\left(y_j\right)\sup_{\theta\in\Theta}\left\|u\left(y_j;\theta\right)\right\| \\
&= O_p\left(\sqrt{n}\right)\max_{1\le i\le n} T_{i,n}\left\|\sum_{j=1}^{n} w_{ij}u\left(y_j;\theta_0\right)\right\| \\
&\quad \sqrt{\frac{1}{n}\sum_{i=1}^{n} T_{i,n}\left\|\lambda_i(\theta_0)\right\|^2}\sqrt{\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}d^2\left(y_j\right)\left[\sup_{\theta\in\Theta}\left\|u\left(y_j;\theta\right)\right\|\right]^2} \\
&= O_p\left(\sqrt{n}\right)\sqrt{o_p\left(\sqrt{\frac{\ln n}{nb_n^{dm+2\varsigma}}}\right)+O\left(\frac{b_n^2}{b_n^{\varsigma}}\right)}\sqrt{o_p\left(\sqrt{\frac{\ln n}{nb_n^{dm+2\varsigma}}}\right)+O\left(\frac{b_n^2}{b_n^{\varsigma}}\right)} \\
&= o_p(1) \qquad\qquad \text{(A.50)}
\end{aligned}
$$

where (A.48), Assumption 2.2.8 (iii), and $\max_{1\le i\le n}\left\|\widehat{V}^{-1}(x_i;\theta_0)\right\| = O_p(1)$ which follows from Lemma A.1.6 and Assumption 2.2.8 (ii) are used in the second equality, Cauchy-Schwartz and Jensen inequalities are used in the third equality, Lemma A.1.2, (A.17) and Lemma A.1.1 are used in the fourth equality, and Assumption 2.2.10 is used in the last step. Last, (A.45), (A.47), (A.49), (A.50), Proposition A.2.3, and the continuity of $I(\theta)$ on $\mathcal{B}_0$ implied by Assumption 2.2.8 yield

$$
n^{1/2}\left(\widehat{\theta}_{LEL}-\theta_0\right) = -I(\theta_0)^{-1} n^{-1/2}A + o_p(1)
$$

which further delivers the desired conclusion by Proposition A.2.7. Q.E.D.

Proof of Theorem 2.2.3: Part (i) can be proved following the same steps in the proof of Theorem 2.2.1 by noticing $u(\cdot)-\widetilde{u}(\cdot) = O_{a.s.}\left(M^{-2}\right)$ from the approximation of the numerical integrals. For part (ii), applying Taylor expansions to both $\nabla_\theta \mathbf{LELL}\left(\widehat{\theta}_{LEL}\right) = 0$ and $\nabla_\theta \mathbf{ALELL}\left(\widehat{\theta}_{ALEL}\right) = 0$ as the first order conditions for (2.35) and (2.36) respectively, we have

$$
n^{1/2}\left(\widehat{\theta}_{LEL}-\widehat{\theta}_{ALEL}\right)
$$

$$
\begin{aligned}
&= -n^{-1/2}\nabla_\theta \mathbf{LELL}(\theta_0)\left[\frac{1}{n}\nabla_{\theta\theta}\mathbf{LELL}(\theta^*)\right]^{-1} + n^{-1/2}\nabla_\theta \mathbf{ALELL}(\theta_0)\left[\frac{1}{n}\nabla_{\theta\theta}\mathbf{ALELL}\left(\widetilde{\theta^*}\right)\right]^{-1} \\
&= \left\{\left[\frac{1}{n}\nabla_{\theta\theta}\mathbf{ALELL}\left(\widetilde{\theta^*}\right)\right]^{-1} - \left[\frac{1}{n}\nabla_{\theta\theta}\mathbf{LELL}(\theta^*)\right]^{-1}\right\} n^{-1/2}\nabla_\theta \mathbf{LELL}(\theta_0) \\
&\quad + \left[\frac{1}{n}\nabla_{\theta\theta}\mathbf{ALELL}\left(\widetilde{\theta^*}\right)\right]^{-1}\left\{n^{-1/2}\nabla_\theta \mathbf{ALELL}(\theta_0) - n^{-1/2}\nabla_\theta \mathbf{LELL}(\theta_0)\right\} \\
&= \frac{1}{n} n^2 O_{a.s.}\left(M^{-2}\right) O_p(1) + O_p(1)\, n^{-1/2} O_{a.s.}\left(M^{-2}\right) \\
&= O_p(nM^{-2}) = o_p(1)
\end{aligned}
$$

where the third equality uses $u(\cdot) - \widetilde{u}(\cdot) = O_{a.s.}\left(M^{-2}\right)$, the fourth uses (A.47)-(A.50), Proposition A.2.3 and Proposition A.2.7, and the last uses the condition $n^{1/2}/M \to 0$. Q.E.D.

APPENDIX B

APPENDIX OF CHAPTER 3

B.1 Testing for Jumps in LIBOR-Swap Yields

The test procedure employed is from Chen and Chapter 2 which is also based on the infinitesimal operator and considers multivariate cases. Here I only describe the test for the univariate case. Suppose we have sample data $\{X_{t=\tau\Delta}\}_{\tau=1}^{n}$ observed over a time span T with the sampling interval Δ and sample size $n = T/\Delta$. We assume $n \to \infty$ and for each Δ, high frequency data are available with the sampling interval $\delta = \Delta/M \to 0$ for integer M. The idea of the test comes from that fact that the diffusion process represented by $dX_t = b(X_t)dt + \sigma(X_t)dW_t$ is the only Markov process with continuous sample paths (Rogers and Williams, 2000). Consequently, if we can reject that the data generating process follows this diffusion model, which is nonparametric, the conclusion can be made that jumps exist in the process.

To test whether LIBOR-Swap yields can be represented by this diffusion process, the infinitesimal operator defined in (3.29) is employed, which is

$$\mathcal{A}f(x) = b(x)f'(x) + \frac{1}{2}\sigma^2(x)f''(x)$$

for the univariate diffusion following (3.29)-(3.30). Then by (3.34) and discussions therein, it is equivalent to test whether $\left(Z_t^x, Z_t^{x^2}\right)'$ is a martingale difference sequence, where

$$Z_t^{\mathbf{x}} = X_t - X_{t-\Delta} - \int_{t-\Delta}^{t} b(X_s)ds$$

$$Z_t^{\mathbf{x}^2} \;=\; X_t^2 - X_{t-\Delta}^2 - \int_{t-\Delta}^{t} [2b(X_s)X_s]\, ds - \int_{t-\Delta}^{t} \sigma^2(X_s)ds \tag{B.1}$$

However, both $b(\cdot)$ and $\sigma^2(\cdot)$ are unkown due to the nonparametric nature of the model. Chen and Chapter 2 suggests estimating $b(\cdot)$ by the following nonparametric local linear estimator

$$\hat{b}(\mathbf{x}) = \sum_{\tau=2}^{n} \hat{W}\left[\frac{X_{(\tau-1)\Delta} - \mathbf{x}}{h}\right] \frac{[X_{\tau\Delta} - \mathbf{X}_{(\tau-1)\Delta}]}{\Delta}, \tag{B.2}$$

where $\hat{W}(\cdot) : \mathbb{R} \to \mathbb{R}$ is a kernel (see Chen and Chapter 2 for the definition) and h is a bandwidth, and estimating $\int_{t-\Delta}^{t} \sigma^2(X_s)ds$ by so-called realized volatility

$$\widehat{[X,X]}_{t-\Delta}^{t} = \sum_{i=1}^{M} \left(X_{t-\Delta+i\delta} - X_{t-\Delta+(i-1)\delta}\right)^2 \tag{B.3}$$

Plugging (B.2) and (B.3) delivers the estimated process $\left(\widehat{Z}_t^x, \widehat{Z}_t^{x^2}\right)'$ of $\left(Z_t^x, Z_t^{x^2}\right)'$. Then a test can be constructed by checking whether $\left(\widehat{Z}_t^x, \widehat{Z}_t^{x^2}\right)'$ is martingale difference sequence, giving us the test statistic $\widehat{M}_0(\overline{p})$ with a preliminary bandwidth $\overline{p}$. Under some regularity conditions, $\widehat{M}_0(\overline{p})$ is shown to have an asymptotic N (0,1) distribution and unitary power by Chen and Chapter 2.

The $\widehat{M}_0(\overline{p})$ test is applied to check whether jumps exist in daily LIBOR-Swap yields with maturities of 3-month, 6-month, 9-month, 2-year, 3-year, 4,-year, 5-year, 7-year and 10-year from August 13, 1990 to December 31, 2008. To calculate the test statistic $\widehat{M}_0(\overline{p})$, the bandwidth parameter h for $\hat{b}(\cdot)$ is chosen as $h = 4\widehat{S}_X n^{-1/4.5}$ where $\widehat{S}_X$ is the standard deviation of the observations. This choice of bandwidth is optimal for the local linear estimation in a mean squared error sense. Since the data is sampled daily, we set $\Delta = 1/52$ and $M = 5$. For the preliminary bandwidth $\overline{p}$, four values (5, 10, 15, 25) are considered for robustness. See Chen and Chapter 2 for details of these choices and other parameters.

Tables B.1 reports the test statistics and p-values of the $\widehat{M}_0(\overline{p})$ test. For yields with all the maturities considered, strong evidence is found for the existence of

Table B.1: **Testing the Existence of Jumps in LIBOR-Swap Yields**

Maturity	$\widehat{M}_0(5)$		$\widehat{M}_0(10)$		$\widehat{M}_0(15)$		$\widehat{M}_0(20)$	
	Test Statistics	P-Values	Test Statistics	P-Values	Test Statistics	P-Values	Test Statistics	P-Values
3-month	48.4519	0.0000	50.3316	0.0000	49.3755	0.0000	49.1913	0.0000
6-month	46.3172	0.0000	48.2866	0.0000	47.2533	0.0000	48.8343	0.0000
9-month	49.9837	0.0000	47.2742	0.0000	48.7761	0.0000	51.1733	0.0000
2-year	49.6132	0.0000	55.3940	0.0000	54.3424	0.0000	53.3277	0.0000
3-year	35.5002	0.0000	38.2101	0.0000	37.1781	0.0000	39.0426	0.0000
4-year	24.1892	0.0000	24.7806	0.0000	24.3199	0.0000	26.3433	0.0000
5-year	23.6966	0.0000	21.6052	0.0000	20.7754	0.0000	20.7048	0.0000
7-year	13.5257	0.0000	12.8414	0.0000	13.6055	0.0000	14.3900	0.0000
10-year	8.6388	0.0000	8.1364	0.0000	8.3683	0.0000	8.7080	0.0000

This table reports results of testing for the existence of jumps in daily LIBOR-Swap yields with maturities of 3-month, 6-month, 9-month, 2-year, 3-year, 4,-year, 5-year, 7-year and 10-year from August 13, 1990 to December 31, 2008. The bandwidth h for the local linear estimator of the drift function is chosen as $4\widehat{S}_{\mathbf{X}}n^{-1/4.5}$, where $\widehat{S}_X$ is the sample standard deviation of the observations. Four values (5, 10, 15, 25) of the preliminary bandwidth $\overline{p}$ are considered for robustness in computing $\widehat{M}_0(\overline{p})$.

jumps: the test statistics range from 8.1364 to 55.3940 and the p-values are all equal to zero. Another observation is that the test statistic tends to be smaller for yields with longer maturities than those with shorter maturities, which is consistent with the fact that short-term interest rates exhibit stronger jump activities due to the impact of monetary polices. Moreover, the test is quite robust to the preliminary bandwidth choice of $\overline{p}$.

B.2 Differences in Swap and Zero-Coupon Yields

As pointed out in Section 3.1.3, there are differences between swap rates and zero-coupon yields: the swap rate is equivalent to a par–bond yield with the coupon rate equal to the swap rate In this section, I construct zero-coupon yields from the LIBOR-Swap rates and compare them to gauge the magnitude of errors in using swap rates as approximations of corresponding zero-coupon yields. The zero-coupon yields $y_{0,t}^{\tau}$ is constructed from the LIBOR-Swap swap rates y_t^{τ} by the following steps: first, a linear interpolation method is applied to obtain swap rates of intermediary maturities, e.g., 6-year, 8-year and 10-year; second, swap rates, treated as yields on coupon-bonds with the coupon rate, are inverted to calculate the zero-coupon yields.

Table B.2 reports the average, in each year from 1990 to 2008, of the daily percentage difference ($\equiv \left(y_t^{\tau} - y_{0,t}^{\tau}\right)/y_{0,t}^{\tau}$) between the swap rate and zero-coupon yield. It can be seen that the errors of using swap rates as approximations of corresponding zero-coupon yields are very small in general: the values range from 4.1606% to -2.0337%. Hence swap rates approximate the corresponding zero-coupon yields fairly well. To further make sure these approximation errors do

Table B.2: **Differences between Swap and Zero-Coupon Yields**

Year	Maturity					
	2-year	3-year	4-year	5-year	7-year	10-year
1990 (%)	4.1606	4.0098	3.8531	3.7391	3.2593	2.9079
1991 (%)	3.2364	3.0531	2.8384	2.5789	2.1926	1.8462
1992 (%)	2.1498	1.9113	1.6041	1.2618	0.6806	-0.1824
1993 (%)	1.8657	1.6905	1.4375	1.1474	0.5876	-0.3528
1994 (%)	2.8536	2.7361	2.5916	2.4353	2.1369	1.6046
1995 (%)	3.1854	3.1046	3.0196	2.9272	2.7408	2.2989
1996 (%)	2.9461	2.8594	2.7566	2.6424	2.3951	1.9189
1997 (%)	3.0936	3.0312	2.9737	2.9047	2.7413	2.4354
1998 (%)	2.8246	2.7693	2.7144	2.6554	2.5172	2.2301
1999 (%)	2.9155	2.8522	2.7955	2.7329	2.5691	2.2577
2000 (%)	3.5133	3.4633	3.4313	3.3915	3.3058	3.1607
2001 (%)	1.9542	1.7992	1.6344	1.4733	1.1465	0.6606
2002 (%)	1.1050	0.9237	0.6996	0.4501	-0.1061	-0.9213
2003 (%)	0.6797	0.4759	0.1720	-0.1861	-0.9154	-2.0337
2004 (%)	1.0652	0.9188	0.7036	0.4491	-0.1215	-1.0405
2005 (%)	2.0394	2.0097	1.9682	1.9087	1.7541	1.4158
2006 (%)	2.7144	2.6916	2.6572	2.6117	2.5182	2.3508
2007 (%)	2.6317	2.5780	2.5064	2.4192	2.2250	1.8926
2008 (%)	1.5328	1.3871	1.2080	1.0267	0.6145	0.0269

This table reports differences between swap rates and zero-coupon yields constructed from LIBOR-Swap rates with maturities of 3-month, 6-month, 9-month, 2-year, 3-year, 4,-year, 5-year, 7-year and 10-year from 08/13, 1990 to 12/31, 2008. In each cell is the yearly average of the daily percentage difference between the swap rate y_t^τ and zero-coupon yield ($y_{0,t}^\tau$) computed as $(y_t^\tau - y_{0,t}^\tau)/y_{0,t}^\tau$.

Table B.3: **Yield Regression Using Constructed Zero-Coupon Yields**

This table reports the results of ”yield regression” in (2.3) using zero-coupon bond yields constructed from daily LIBOR-Swap rates with maturities as indicated in the table from August 13, 1990 to December 31, 2008. The 3-month LIBOR rate is used as the spot rate r_t. In the ”s.e.” row are the Newey-West standard errors of $\phi_{\tau T}$.

	$y_{t+1}^{(\tau-1)} - y_t^{\tau} = \text{constant}+\phi_{\tau T}\left(y_t^{\tau} - r_t\right) / (\tau-1) + residual$							
Maturity	6-month	9-month	2-year	3-year	4-year	5-year	7-year	10-year
ϕ_{nT}	0.0096	0.0090	0.0025	-0.0001	-0.0013	-0.0026	-0.0039	-0.0051
$s.e.$	0.0027	0.0026	0.0030	0.0036	0.0041	0.0008	0.0054	0.0027

not affect results in Table 3.2 of testing the ”expectation hypothesis”, estimates of slope coefficients $\phi_{\tau T}$ in the ”yield regression” of (3.3) are computed and reported in Table B.3, using constructed zero-coupon yields with maturities of 2-year, 3-year, 4,-year, 5-year, 7-year and 10-year and LIBOR rates of 3-month, 6-month, and 9-month. Comparing Table B.3 with Table 3.2, we see that differences in regression coefficients estimates are extremely small and can in fact be neglected. Therefore, swap rates are good approximations to the zero-coupon yields as well for testing the ”expectation hypothesis”.

APPENDIX C

APPENDIX OF CHAPTER 4

Throughout the Appendix, let $g_i(\tau,\theta)$, $g_{ij}(\tau,\theta)$, $Z^i_\tau(\theta)$, and $Z^{ij}_\tau(\theta)$ be defined as in (4.38)-(4.51). I let $M_0(p)$ be defined in the same way as $\widehat{M}_0(p)$ in (4.28) with the unobservable sample $\{Z_\tau = Z_{\tau\Delta}(\theta_0)\}_{\tau=1}^n$, where $\theta_0 = p\lim\widehat{\theta}$, replacing the estimated processes samples $\{\widehat{Z}_\tau = Z_{\tau\Delta}(\widehat{\theta})\}_{\tau=1}^n$. Also, $C \in (1,\infty)$ denotes a generic bounded constant.

Proof of Theorem 4.4.1. See ChV.19-20 of Rogers and Williams(2000), or Theorem 21.7 of Kallengber(2002), or Proposition 2.4 of ChVII in Revuz and Yor(2005). Q.E.D.

Proof of Theorem 4.1.2. See Proposition 4.6 of Karatzas and Shreve(1991, Ch5.4) Q.E.D.

Proof of Theorem 4.4.1. It suffices to show Theorems C.1-C.3 below. Theorem C.1 implies that replacing $\{Z_\tau\}_{\tau=1}^n$ by $\{\widehat{Z}_\tau\}_{\tau=1}^n$ has no impact on the limit distribution of $\widehat{M}_0(p)$; Theorem C.2 says that the use of truncated process $\{Z_{q,\tau}\}_{\tau=1}^n$ rather than the original $\{Z_\tau\}_{\tau=1}^n$ does not affect the limit distribution of $\widehat{M}_0(p)$ for q sufficiently large. The assumption that $Z_{q,\tau}$ is independent of $\{Z_{\tau-m}\}_{m=q+1}^\infty$ when q is large greatly simplifies the derivation of asymptotic normality of $\widehat{M}_0(p)$. Q.E.D.

Theorem C.1. Under the conditions of Theorem 4.4.1, $\widehat{M}_0(p) - M_0(p) \to^p 0$.

Theorem C.2. Let $M_{0q}(p)$ be defined as $M_0(p)$ with $\{Z_{q,\tau}\}_{\tau=1}^n$ replacing $\{Z_\tau\}_{\tau=1}^n$,where $\{Z_{q,\tau}\}$ is as in Assumption 4.4.2. Then under the conditions of Theorem 4.4.1 and $q = p^{1+\frac{1}{4b-2}}\left(\ln^2 n\right)^{\frac{1}{2b-1}}$, $M_{0q}(p) - M_0(p) \to^p 0$.

Theorem C.3. Under the conditions of Theorem 4.4.1 and $q = p^{1+\frac{1}{4b-2}}\left(\ln^2 n\right)^{\frac{1}{2b-1}}$, $M_{0q}(p) \to^d N(0,1)$.

Proof of Theorem C.1. Note that $Z_\tau(\theta)$ has components $Z_\tau^i(\theta) = X_{\tau\Delta}^i - X_{(\tau-1)\Delta}^i + g_i(\tau,\theta)$ and $Z_\tau^{ij}(\theta) = X_{\tau\Delta}^i X_{\tau\Delta}^j - X_{(\tau-1)\Delta}^i X_{(\tau-1)\Delta}^j + g_{ij}(\tau,\theta)$ and similarly $\widehat{Z}_\tau$ has components $\widehat{Z}_\tau^i = X_{\tau\Delta}^i - X_{(\tau-1)\Delta}^i + g_i(\tau,\widehat{\theta})$ and $\widehat{Z}_\tau^{ij} = X_{\tau\Delta}^i X_{\tau\Delta}^j - X_{(\tau-1)\Delta}^i X_{(\tau-1)\Delta}^j + g_{ij}(\tau,\widehat{\theta})$ respectively. By the mean value theorem, we have $\widehat{Z}_\tau^i = Z_\tau^i - g_i^{'}(\tau,\overline{\theta})^{'}(\widehat{\theta}-\theta_0)$ for some $\overline{\theta}$ between $\widehat{\theta}$ and θ_0, where $g_i^{'}(\tau,\theta) = \frac{\partial}{\partial\theta} g_i(\tau,\theta)$. By the Cauchy-Schwartz inequality and Assumptions 4.4.3-4.4.4,

$$\sum_{\tau=1}^{n}\left(\widehat{Z}_\tau^i - Z_\tau^i\right)^2 \leqq n\left\|\widehat{\theta}-\theta_0\right\|^2 n^{-1}\sum_{\tau=1}^{n}\sup_{\theta\in\Theta_0}\left\|g_i'(\tau,\theta)\right\|^2 = O_p(1) \tag{C.1}$$

for any $i = 1,\cdots,d$, where Θ_0 is a neighborhood of θ_0. By similar reasoning, we have

$$\sum_{\tau=1}^{n}\left(\widehat{Z}_\tau^{ij} - Z_\tau^{ij}\right)^2 \leqq n\left\|\widehat{\theta}-\theta_0\right\|^2 n^{-1}\sum_{\tau=1}^{n}\sup_{\theta\in\Theta_0}\left\|g_i'(\tau,\theta)\right\|^2 = O_p(1) \tag{C.2}$$

for any $i, j = 1,\cdots,d$. (C.1) and (C.2) together imply that

$$\begin{aligned}
&\sum_{\tau=1}^{n}\left\|\widehat{Z}_\tau - Z_\tau\right\|^2 \\
= &\sum_{\tau=1}^{n}\left[\sum_{i=1}^{d}\left(\widehat{Z}_\tau^i - Z_\tau^i\right)^2 + \sum_{i=1}^{d}\sum_{j=1}^{d}\left(\widehat{Z}_\tau^{ij} - Z_\tau^{ij}\right)^2\right] \\
= &\sum_{i=1}^{d}\left[\sum_{\tau=1}^{n}\left(\widehat{Z}_\tau^i - Z_\tau^i\right)^2\right] + \sum_{i=1}^{d}\sum_{j=1}^{d}\left[\sum_{\tau=1}^{n}\left(\widehat{Z}_\tau^{ij} - Z_\tau^{ij}\right)^2\right] \\
= &O_p(1)
\end{aligned} \tag{C.3}$$

Now put $n_m = n - |m|$, and let $\widetilde{\sigma}_m^{(1,0)}(0,v)$ be defined in the same way as $\widehat{\sigma}_m^{(1,0)}(0,v)$ in (4.28), with $\{Z_\tau\}_{\tau=1}^n$ replacing $\{\widehat{Z}_\tau\}_{\tau=1}^n$. To show $\widehat{M}_0(p) - M_0(p) \to^p 0$, it is sufficient to prove

$$\widehat{D}_0^{-\frac{1}{2}}(p)\int\sum_{m=1}^{n-1}k^2(m/p)n_m\left[\left\|\widehat{\sigma}_m^{(1,0)}(0,v)\right\|^2 - \left\|\widetilde{\sigma}_m^{(1,0)}(0,v)\right\|^2\right]dW(v) \to^p 0 \tag{C.4}$$

$\widehat{C}_0(p) - \widetilde{C}_0(p) = O_p(n^{-1/2})$, and $\widehat{D}_0(p) - \widetilde{D}_0(p) = o_p(1)$, where $\widetilde{C}_0(p)$ and $\widetilde{D}_0(p)$ are defined in the same way as $\widehat{C}_0(p)$ and $\widehat{D}_0(p)$ in (4.32) with $\{Z_\tau\}_{\tau=1}^n$ replacing $\{\widehat{Z}_\tau\}_{\tau=1}^n$. To save space, I focus on the proof of (C.4); the proofs for $\widehat{C}_0(p) - \widetilde{C}_0(p) = O_p(n^{-1/2})$, and $\widehat{D}_0(p) - \widetilde{D}_0(p) = o_p(1)$ are routine. Note that it is necessary to achieve the convergence rate $O_p(n^{-1/2})$ for $\widehat{C}_0(p) - \widetilde{C}_0(p)$ to make sure that replacing $\widehat{C}_0(p)$ with $\widetilde{C}_0(p)$ has asymptotically negligible impact given $p/n \to 0$.

Since

$$\begin{aligned} & \left\|\widehat{\sigma}_m^{(1,0)}(0,v)\right\|^2 - \left\|\widetilde{\sigma}_m^{(1,0)}(0,v)\right\|^2 \\ = & \textstyle\sum_{i=1}^d \left[\left|\widehat{\sigma}_{m,i}^{(1,0)}(0,v) - \widetilde{\sigma}_{m,i}^{(1,0)}(0,v)\right|^2\right] + \sum_{i=1}^d \sum_{j=1}^d \left[\left|\widehat{\sigma}_{m,ij}^{(1,0)}(0,v) - \widetilde{\sigma}_{m,ij}^{(1,0)}(0,v)\right|^2\right] \end{aligned} \tag{C.5}$$

where $\widehat{\sigma}_{m,i}^{(1,0)}(0,v)$ and $\widehat{\sigma}_{m,ij}^{(1,0)}(0,v)$ for $i,j = 1,\cdots,d$ are the components of $\widehat{\sigma}_m^{(1,0)}(0,v)$ and correspondingly $\widetilde{\sigma}_{m,i}^{(1,0)}(0,v)$ and $\widetilde{\sigma}_{m,ij}^{(1,0)}(0,v)$ for $i,j = 1,\cdots,d$ are the components of $\widetilde{\sigma}_m^{(1,0)}(0,v)$. By (C.5), it is sufficient for (C.4) to show that

$$\widehat{D}_0^{-\frac{1}{2}}(p) \int \sum_{m=1}^{n-1} k^2(m/p) n_m \left[\left|\widehat{\sigma}_{m,i}^{(1,0)}(0,v)\right|^2 - \left|\widetilde{\sigma}_{m,i}^{(1,0)}(0,v)\right|^2\right] dW(v) \to^p 0, \tag{C.6}$$

for $i = 1,\cdots,d$, and

$$\widehat{D}_0^{-\frac{1}{2}}(p) \int \sum_{m=1}^{n-1} k^2(m/p) n_m \left[\left|\widehat{\sigma}_{m,ij}^{(1,0)}(0,v)\right|^2 - \left|\widetilde{\sigma}_{m,ij}^{(1,0)}(0,v)\right|^2\right] dW(v) \to^p 0, \tag{C.7}$$

for $i,j = 1,\cdots,d$. We will only show (C.6) here and the proof of (C.7) is similar.

To show (C.6), I first decompose

$$\int \sum_{m=1}^{n-1} k^2(m/p) n_m \left[\left|\widehat{\sigma}_{m,i}^{(1,0)}(0,v)\right|^2 - \left|\widetilde{\sigma}_{m,i}^{(1,0)}(0,v)\right|^2\right] dW(v) = \widehat{A}_1 + 2Re(\widehat{A}_2) \tag{C.8}$$

where

$$\widehat{A}_1 = \int \sum_{m=1}^{n-1} k^2(m/p) n_m \left|\widehat{\sigma}_{m,i}^{(1,0)}(0,v) - \widetilde{\sigma}_{m,i}^{(1,0)}(0,v)\right|^2 dW(v)$$

$$\widehat{A_2} = \int \sum_{m=1}^{n-1} k^2(m/p) n_m \left[\widehat{\sigma}_{m,i}^{(1,0)}(0,v) - \widetilde{\sigma}_{m,i}^{(1,0)}(0,v)\right] \widetilde{\sigma}_{m,i}^{(1,0)}(0,v)^* dW(v)$$

where $Re(\widehat{A_2})$ denote the real part of $\widehat{A_2}$ and $\widetilde{\sigma}_{m,i}^{(1,0)}(0,v)^*$ denote the complex conjugate of $\widetilde{\sigma}_{m,i}^{(1,0)}(0,v)$. Then (C.6) follows from the following Propositions C.1 and C.2 and $p \to \infty$ as $n \to \infty$. Q.E.D.

Proposition C.1. Under the conditions of Theorem 4.4.1, $\widehat{A_1} = O_p(1)$.

Proposition C.2. Under the conditions of Theorem 4.4.1, $p^{-1/2}\widehat{A_2} = o_p(1)$.

Proof of Proposition C.1. Put $\widehat{\delta}_\tau(v) = e^{iv'\widehat{Z}_\tau} - e^{iv'Z_\tau}$ and $\psi_\tau(v) = e^{iv'Z_\tau} - \varphi(v)$, where $\varphi(v) = Ee^{iv'Z_\tau}$. Then straightforward algebra yields that for $m > 0$,

$$\begin{aligned}
& \widehat{\sigma}_{m,i}^{(1,0)}(0,v) - \widetilde{\sigma}_{m,i}^{(1,0)}(0,v) \\
= \;& in_m^{-1} \sum_{\tau=m+1}^{n} (\widehat{Z}_{\tau,i} - Z_{\tau,i}) \widehat{\delta}_{\tau-m}(v) \\
& -i \left[n_m^{-1} \sum_{\tau=m+1}^{n} (\widehat{Z}_{\tau,i} - Z_{\tau,i}) \right] \left[n_m^{-1} \sum_{\tau=m+1}^{n} \widehat{\delta}_{\tau-m}(v) \right] \\
& +in_m^{-1} \sum_{\tau=m+1}^{n} Z_{\tau,i} \widehat{\delta}_{\tau-m}(v) - i \left[n_m^{-1} \sum_{\tau=m+1}^{n} Z_{\tau,i} \right] \left[n_m^{-1} \sum_{\tau=m+1}^{n} \widehat{\delta}_{\tau-m}(v) \right] \\
& +in_m^{-1} \sum_{\tau=m+1}^{n} (\widehat{Z}_{\tau,i} - Z_{\tau,i}) \psi_{\tau-m}(v) \\
& -i \left[n_m^{-1} \sum_{\tau=m+1}^{n} (\widehat{Z}_{\tau,i} - Z_{\tau,i}) \right] \left[n_m^{-1} \sum_{\tau=m+1}^{n} \psi_{\tau-m}(v) \right] \\
= \;& i \left[\widehat{B}_{1m}(v) - \widehat{B}_{2m}(v) + \widehat{B}_{3m}(v) - \widehat{B}_{4m}(v) + \widehat{B}_{5m}(v) - \widehat{B}_{6m}(v) \right]
\end{aligned} \tag{C.9}$$

Then it follows that $\widehat{A_1} \leqq 8 \sum_{a'=1}^{6} \sum_{m=1}^{n-1} k^2(m/p) n_m \int \left| \widehat{B}_{a'm}(v) \right|^2 dW(v)$. Proposition C.1 follows from Lemmas C.1-C.6 below and $p/n \to 0$. Q.E.D.

Lemma C.1. $\sum_{m=1}^{n-1} k^2(m/p) n_m \int \left| \widehat{B}_{1m}(v) \right|^2 dW(v) = O_p(p/n)$.

Lemma C.2. $\sum_{m=1}^{n-1} k^2(m/p) n_m \int \left| \widehat{B}_{2m}(v) \right|^2 dW(v) = O_p(p/n)$.

Lemma C.3. $\sum_{m=1}^{n-1} k^2(m/p)n_m \int \left|\widehat{B}_{3m}(v)\right|^2 dW(v) = O_p(p/n)$.

Lemma C.4. $\sum_{m=1}^{n-1} k^2(m/p)n_m \int \left|\widehat{B}_{4m}(v)\right|^2 dW(v) = O_p(p/n)$.

Lemma C.5. $\sum_{m=1}^{n-1} k^2(m/p)n_m \int \left|\widehat{B}_{5m}(v)\right|^2 dW(v) = O_p(1)$.

Lemma C.6. $\sum_{m=1}^{n-1} k^2(m/p)n_m \int \left|\widehat{B}_{6m}(v)\right|^2 dW(v) = O_p(p/n)$.

Now let $a_n(m) = n_m^{-1}k^2(m/p)$. In the following, I will show these lemmas above.

Proof of Lemma C.1. By the Cauchy-Schwartz inequality and the inequality that $\left|e^{iz_1} - e^{iz_2}\right| \leqq |z_1 - z_2|$ for any real-valued variables z_1 and z_2, I have

$$\begin{aligned} \left|\widehat{B}_{1m}(v)\right|^2 &\leqq \left[n_m^{-1}\sum_{\tau=1}^{n}(\widehat{Z}_{\tau,i} - Z_{\tau,i})^2\right]\left[n_m^{-1}\sum_{\tau=1}^{n}\left|\widehat{\delta}_\tau(v)\right|^2\right] \\ &\leqq \|v\|^2\left[n_m^{-1}\sum_{\tau=1}^{n}(\widehat{Z}_{\tau,i} - Z_{\tau,i})^2\right]^2 \end{aligned}$$

It follows from (C.1) and Assumptions 4.4.5-4.4.6 that

$$\begin{aligned} &\int\sum_{m=1}^{n-1}k^2(m/p)n_m^{-1}\left|\widehat{B}_{1m}(v)\right|^2 dW \\ \leqq\; &\left[\sum_{m=1}^{n-1}a_n(m)\right]\left[\sum_{\tau=1}^{n}(\widehat{Z}_{\tau,i} - Z_{\tau,i})^2\right]^2\int\|v\|^2 dW(v) = O_p(p/n) \end{aligned} \tag{C.10}$$

where I made use of the fact that

$$\sum_{m=1}^{n-1}a_n(m) = \sum_{m=1}^{n-1}n_m^{-1}k^2(m/p) = O(p/n) \tag{C.11}$$

given $p = cn^\lambda$ for $\lambda \in (0, 1/2)$, as shown in Hong(1999, A.15, Page 1213). Q.E.D.

Proof of Lemma C.2. By the inequality that $\left|e^{iz_1} - e^{iz_2}\right| \leqq |z_1 - z_2|$ for any real-valued variables z_1 and z_2, I have

$$\begin{aligned}\left|\widehat{B}_{2m}(v)\right|^2 &\leqq \left[n_m^{-1}\sum_{\tau=1}^{n}\left|\widehat{Z}_{\tau,i} - Z_{\tau,i}\right|\right]^2\left[n_m^{-1}\sum_{\tau=1}^{n}\left|v\widehat{Z}_{\tau,i} - vZ_{\tau,i}\right|\right]^2 \\ &\leqq \|v\|^2\left[n_m^{-1}\sum_{\tau=1}^{n}(\widehat{Z}_{\tau,i} - Z_{\tau,i})^2\right]^2\end{aligned}$$

By the same reasoning as that of Lemma C.1, the desired result follows. Q.E.D.

Proof of Lemma C.3. Using the inequality that $\left|e^{iz} - 1 - iz\right| \leqq |z|^2$ for any real-valued variables z, I have

$$\begin{aligned}&\left|e^{iv'\widehat{Z}_{\tau-m}} - e^{iv'Z_{\tau-m}} - iv\left[\widehat{Z}_{\tau-m,i} - Z_{\tau-m,i}\right]e^{iv'Z_{\tau-m}}\right| \\ \leqq\ &\|v\|^2\left[\widehat{Z}_{\tau-m,i} - Z_{\tau-m,i}\right]^2 \end{aligned} \tag{C.12}$$

A second order Taylor series expansion yields

$$\widehat{Z}_{\tau-m,i} = Z_{\tau-m,i} - (g_i'(\tau - m, \theta_0))'(\widehat{\theta} - \theta_0) - \frac{1}{2}(\widehat{\theta} - \theta_0)'g_i''(\tau - m, \overline{\theta})(\widehat{\theta} - \theta_0) \tag{C.13}$$

for some $\overline{\theta}$ between $\widehat{\theta}$ and θ_0, where $g_i''(\tau, \theta) \equiv \frac{\partial^2}{\partial\theta\partial\theta'}g(\tau, \theta)$. Put $\xi_\tau(v) = g_i'(\tau, \theta_0)e^{iv'Z_\tau}$. Then (C.12) and (C.3) imply that

$$\begin{aligned}&\left|e^{iv'\widehat{Z}_{\tau-m}} - e^{iv'Z_{\tau-m}} - iv\xi_{\tau-m}(v)(\widehat{\theta} - \theta_0)\right| \\ \leqq\ &\|v\|^2\left[\widehat{Z}_{\tau-m,i} - Z_{\tau-m,i}\right]^2 + \|v\|\left\|\widehat{\theta} - \theta_0\right\|^2\sup_{\theta\in\Theta_0}\left\|g_i''(\tau - m, \theta)\right\|\end{aligned}$$

where Θ_0 is a neighborhood of θ_0.

Henceforth, by (C.9), I obtain

$$
\begin{aligned}
& n_m \left|\widehat{B}_{3m}(v)\right| \\
\leqq\ & \|v\| \left\|\widehat{\theta}-\theta_0\right\| \left\| \sum_{\tau=m+1}^{n} Z_{\tau,i}\xi_{\tau-m}(v) \right\| + \|v\|^2 \sum_{\tau=m+1}^{n} \left|Z_{\tau,i}\right| \left[\widehat{Z}_{\tau-m,i} - Z_{\tau-m,i}\right]^2 \\
& + \|v\| \left\|\widehat{\theta}-\theta_0\right\|^2 \sum_{\tau=m+1}^{n} \left|Z_{\tau,i}\right| \sup_{\theta\in\Theta_0} \left\|g_i''(\tau-m,\theta)\right\|
\end{aligned}
$$

Then it follows from Assumptions 4.4.1-4.47 and (C.11) that

$$
\begin{aligned}
& \sum_{m=1}^{n-1} \int k^2(m/p) n_m \left|\widehat{B}_{3m}(v)\right|^2 dW(v) \\
\leqq\ & 4 \left\| \sqrt{n}(\widehat{\theta}-\theta_0)\right\|^2 \sum_{m=1}^{n-1} k^2(m/p) \int \left\| n_m^{-1} \sum_{\tau=m+1}^{n} Z_{\tau,i}\xi_{\tau-m}(v)\right\|^2 \|v\|^2\, dW(v) \\
& +4 \left\| \sqrt{n}(\widehat{\theta}-\theta_0)\right\|^4 \left(n^{-1}\sum_{\tau=1}^{n} Z_{\tau,i}^2\right) \left\{ n^{-1} \sum_{\tau=1}^{n} \left[\sup_{\theta\in\Theta_0} \left\|g_i'(\tau,\theta)\right\|\right]^4 \right\} \\
& \times \left[\sum_{m=1}^{n-1} a_n(m)\right] \int \|v\|^4\, dW(v) \\
& +4 \left\| \sqrt{n}(\widehat{\theta}-\theta_0)\right\|^4 \left(n^{-1}\sum_{\tau=1}^{n} Z_{\tau,i}^2\right) \left\{ n^{-1} \sum_{\tau=1}^{n} \left[\sup_{\theta\in\Theta_0} \left\|g_i''(\tau,\theta)\right\|\right]^2 \right\} \\
& \times \left[\sum_{m=1}^{n-1} a_n(m)\right] \int \|v\|^2\, dW(v) \\
=\ & O_p(p/n) \qquad\qquad (C.14)
\end{aligned}
$$

by the fact that $E\left\|\sum_{\tau=m+1}^{n} Z_{\tau,i}\xi_{\tau-m}(v)\right\|^2 \leqq Cn_m$ given $E\left(Z_{\tau,i} \mid I_{\tau-1}\right) = 0\ a.s.$ under H_0 and Assumptions 4.4.1 and 4.4.3. Q.E.D.

Proof of Lemma C.4. By the Cauchy-Schwartz inequality,

$$
\left|\widehat{B}_{4m}(v)\right|^2 \leqq \left(n_m^{-1} \sum_{\tau=m+1}^{n} Z_{\tau,i}\right)^2 n_m^{-1} \sum_{\tau=m+1}^{n} \left|\widehat{\delta}_\tau(v)\right|
$$

Then by this inequality, Cauchy-Schwartz again, and $\left|\widehat{\delta}_\tau(v)\right| \leqq \left|v'\left(\widehat{Z}_\tau - Z_\tau\right)\right|$,

$$
\begin{aligned}
& \sum_{m=1}^{n-1} \int k^2(m/p) n_m \left|\widehat{B}_{4m}(v)\right|^2 dW(v) \\
\leqq & \sum_{m=1}^{n-1} k^2(m/p) \left(n_m^{-1} \sum_{\tau=m+1}^{n} Z_{\tau,i}\right)^2 \left[\sum_{\tau=1}^{n} \left\|\widehat{Z}_\tau - Z_\tau\right\|^2\right] \int \|v\|^2 dW(v) \\
= & O_p(p/n)
\end{aligned}
$$

given (C.3) and (C.11), and $E\left(\sum_{\tau=m+1}^{n} Z_{\tau,i}\right)^2 = \sigma^2 n_m$ by H_0, the *m.d.s.* hypothesis of $\{Z_\tau\}$. Q.E.D.

Proof of Lemma C.5. By the second order Taylor series expansion in (C.13),

$$
\begin{aligned}
-\widehat{B}_{5m}(v) &= (\widehat{\theta} - \theta_0)' n_m^{-1} \sum_{\tau=m+1}^{n} g_i'(\tau, \theta_0) \psi_{\tau-m}(v) \\
&\quad + \frac{1}{2}(\widehat{\theta} - \theta_0)' \left[n_m^{-1} \sum_{\tau=m+1}^{n} g_i''(\tau, \overline{\theta}) \psi_{\tau-m}(v)\right] (\widehat{\theta} - \theta_0)
\end{aligned}
$$

for some $\overline{\theta}$ between $\widehat{\theta}$ and θ_0. Then I have

$$
\begin{aligned}
& \sum_{m=1}^{n-1} k^2(m/p) n_m \int \left|\widehat{B}_{5m}(v)\right|^2 dW(v) \\
\leqq & 2\left\|\sqrt{n}(\widehat{\theta} - \theta_0)\right\|^2 \sum_{m=1}^{n-1} k^2(m/p) \int \left\|n_m^{-1} \sum_{\tau=m+1}^{n} g_i'(\tau, \theta_0) \psi_{\tau-m}(v)\right\|^2 dW(v) \\
& +2\left\|\sqrt{n}(\widehat{\theta} - \theta_0)\right\|^4 \left[n^{-1} \sum_{\tau=1}^{n} \sup_{\theta \in \Theta_0} \left\|g_i''(\tau, \theta)\right\|\right]^2 \left[\sum_{m=1}^{n-1} a_n(m)\right] \int dW(v) \\
= & O_p(1) + O_p(p/T) \qquad \text{(C.15)}
\end{aligned}
$$

where the last term is $O_p(p/T)$ given (C.11) and the first term is $O_p(1)$, as is shown below:

Put $\eta_m(v) = E\left(g_i'(\tau,\theta_0)\psi_{\tau-m}(v)\right) = Cov[g_i'(\tau,\theta_0), \psi_{\tau-m}(v)]$. Then

$$\sup_{v\in\mathbb{R}^{d'}} \sum_{m=1}^{\infty} \|\eta_m(v)\| \leqq C$$

by Assumption 4.4.7. Then expressing the moments in terms of cumulants by the well-known formulas(see Hannan, 1970, (5.1), Page 23 for real-valued processes and also Stratonovich(1963), chapter 1 and Leonov and Shiryaev (1959) for more details), I can obtain

$$\begin{aligned}
& n_m E\left\|n_m^{-1}\sum_{\tau=m+1}^{n} g_i'(\tau,\theta_0)\psi_{\tau-m}(v) - \eta_m(v)\right\|^2 \\
\leqq\ & \sum_{r=-n_m}^{n_m} \left\|Cov[g_i'(\tau,\theta_0), g_i'(-r,\theta_0)']\right\| \cdot |\sigma_r(v,-v)| \\
& + \sum_{r=-n_m}^{n_m} \left\|\eta_{m+|r|}(-v)\right\| \cdot \left\|\eta_{m-|r|}(v)\right\| + \sum_{r=-n_m}^{n_m} \left\|\kappa_{m,|r|,m+|r|}(v)\right\| \\
\leqq\ & C
\end{aligned} \tag{C.16}$$

given Assumption 4.4.7, where $\kappa_{m,l,r}(v)$ is as in Assumption 4.4.7. As a result, from (C.11) and (C.16), $|k(\cdot)| \leqq 1$, and $p/n \to 0$, I get

$$\begin{aligned}
& \sum_{m=1}^{n-1} k^2(m/p) E \int \left\|n_m^{-1}\sum_{\tau=m+1}^{n} g_i'(\tau,\theta_0)\psi_{\tau-m}(v)\right\|^2 dW(v) \\
\leqq\ & C\sum_{m=1}^{n-1}\int \|\eta_m(v)\|^2\, dW(v) + C\sum_{m=1}^{n-1} a_n(m) \\
=\ & O(1) + O(p/n) = O(1)
\end{aligned}$$

Therefore the first term in (C.10) is $O_p(1)$. Q.E.D.

Proof of Lemma C.6. The proof is analogous to that of Lemma C.4. Q.E.D.

Proof of Proposition C.2. Given the decomposition in (C.9), I have

$$\left\|\left[\widehat{\sigma}_{m,i}^{(1,0)}(0,v)-\widetilde{\sigma}_{m,i}^{(1,0)}(0,v)\right]\widetilde{\sigma}_{m,i}^{(1,0)}(0,v)^{*}\right| \leqq \sum_{a'=1}^{6}\left|\widehat{B}_{a'm}(v)\right|\left|\widetilde{\sigma}_{m,i}^{(1,0)}(0,v)\right| \tag{C.17}$$

where $\widehat{B}_{a'm}(v)$ is defined in (C.9). By the Cauchy-Schwartz inequality,

$$\begin{aligned}
& \sum_{m=1}^{n-1}k^2(m/p)n_m\int\left|\widehat{B}_{a'm}(v)\right|\cdot\left|\widetilde{\sigma}_{m,i}^{(1,0)}(0,v)\right|dW(v)\\
\leqq\; & \left[\sum_{m=1}^{n-1}k^2(m/p)n_m\int\left|\widehat{B}_{a'm}(v)\right|^2 dW(v)\right]^{1/2}\\
& \times\left[\sum_{m=1}^{n-1}k^2(m/p)n_m\int\left|\widetilde{\sigma}_{m,i}^{(1,0)}(0,v)\right|^2 dW(v)\right]^{1/2}\\
=\; & O_p(p^{1/2}/n^{1/2})O_p(p^{1/2}) = o_p(p^{1/2}), a' = 1,2,3,4,6,
\end{aligned}$$

given Lemmas C.1-C.4 and C.6, and $p/n\to 0$, where

$$p^{-1}\sum_{m=1}^{n-1}k^2(m/p)n_m\int\left|\widetilde{\sigma}_{m,i}^{(1,0)}(0,v)\right|^2 dW(v) = O_p(1)$$

by Markov's inequality, the *m.d.s.*property of $\{Z_\tau\}$ under H_0, and (C.9).

Then consider the case $a' = 5$. By Assumptions 4.4.1-4.4.7,

$$\begin{aligned}
& \sum_{m=1}^{n-1}k^2(m/p)n_m\int\left|\widehat{B}_{5m}(v)\right|\cdot\left|\widetilde{\sigma}_{m,i}^{(1,0)}(0,v)\right|dW(v)\\
\leqq\; & \left\|\widehat{\theta}-\theta_0\right\|\sum_{m=1}^{n-1}k^2(m/p)n_m\int\left\|n_m^{-1}\sum_{\tau=m+1}^{n}g_i'(\tau,\theta_0)\psi_{\tau-m}(v)\right\|\left|\widetilde{\sigma}_{m,i}^{(1,0)}(0,v)\right|dW(v)\\
& +n\left\|\widehat{\theta}-\theta_0\right\|^2\left[n^{-1}\sum_{m=1}^{n-1}\sup_{\theta\in\Theta_0}\left\|g_i''(\tau,\theta)\right\|\right]\sum_{m=1}^{n-1}k^2(m/p)\int\left|\widetilde{\sigma}_{m,i}^{(1,0)}(0,v)\right|dW(v)\\
=\; & O_p(1+p/n^{1/2})+O_p(p/n^{1/2}) = o_p(p^{1/2})
\end{aligned} \tag{C.18}$$

given $p \to \infty$ and $p/n \to 0$, where I have used the fact that

$$n_m E \left|\widetilde{\sigma}_{m,i}^{(1,0)}(0,v)\right|^2 \leq C$$

by the *m.d.s.* property of $\{Z_\tau\}$ under H_0 and the fact that the first term in (C.18) is $O_p(1 + p/n^{1/2})$, as shown below:

By (C.16) and Cauchy-Schwartz inequality, I have

$$\begin{aligned}
& E\left[\left\| n_m^{-1} \sum_{\tau=m+1}^{n} g_i'(\tau,\theta_0)\psi_{\tau-m}(v)\right\| \left|\widetilde{\sigma}_{m,i}^{(1,0)}(0,v)\right|\right] \\
\leqq & \left[E\left\| n_m^{-1} \sum_{\tau=m+1}^{n} g_i'(\tau,\theta_0)\psi_{\tau-m}(v)\right\|^2\right]^{1/2} \left[E\left|\widetilde{\sigma}_{m,i}^{(1,0)}(0,v)\right|^2\right]^{1/2} \\
\leqq & C\left[\|\eta_m(v)\| + Cn_m^{-1/2}\right] n_m^{-1/2}
\end{aligned}$$

and consequently

$$\begin{aligned}
& n^{-\frac{1}{2}} \sum_{m=1}^{n-1} k^2(m/p) n_m E \int \left\| n_m^{-1} \sum_{\tau=m+1}^{n} g_i'(\tau,\theta_0)\psi_{\tau-m}(v)\right\| \left|\widetilde{\sigma}_{m,i}^{(1,0)}(0,v)\right| dW(v) \\
\leqq & C \sum_{m=1}^{n-1} \int \|\eta_m(v)\| \, dW(v) + Cn^{-\frac{1}{2}} \sum_{m=1}^{n-1} k^2(m/p) \\
= & O(1 + p/n^{1/2})
\end{aligned}$$

given $|k(\cdot)| \leqq 1$ and Assumption 4.4.7. Q.E.D.

Proof of Theorem C.2. The proof is similar to Theorem C.1. By the same reasoning as that of (C.4)-(C.7), we will consider only the case $i = 1 \cdots, d$. Let $\widehat{A}_{1,q}$ and $\widehat{A}_{2,q}$ be defined in the same way as $\widehat{A}_1$ and $\widehat{A}_2$ in (C.8), with $\{Z_{q,\tau}\}_{\tau=1}^{n}$ replacing $\{\widehat{Z}_\tau\}_{\tau=1}^{n}$. It is enough to show that $p^{-\frac{1}{2}}\widehat{A}_{1,q} \to^p 0$ and $p^{-\frac{1}{2}}\widehat{A}_{2,q} \to^p 0$.

Let $\delta_{q,\tau} = e^{iv'Z_\tau} - e^{iv'Z_{q,\tau}}$ and $\psi_{q,\tau}(v) = e^{iv'Z_{q,\tau}} - \varphi_q(v)$, where $\varphi_q(v) = E[e^{iv'Z_{q,\tau}}]$. Let $\widetilde{\sigma}^{(1,0)}_{q,m}(0,v)$ be defined as $\widetilde{\sigma}^{(1,0)}_{m}(0,v)$ with $\{Z_{q,\tau}\}_{\tau=1}^n$ replacing $\{Z_\tau\}_{\tau=1}^n$. Then similar to (C.9), I have

$$
\begin{aligned}
& \widetilde{\sigma}^{(1,0)}_{m,i}(0,v) - \widetilde{\sigma}^{(1,0)}_{q,m,i}(0,v) \\
= \ & in_m^{-1}\sum_{\tau=m+1}^{n}(Z_{\tau,i} - Z_{q,\tau,i})\delta_{q,\tau-m}(v) - i\left[n_m^{-1}\sum_{\tau=m+1}^{n}(Z_{\tau,i} - Z_{q,\tau,i})\right]\left[n_m^{-1}\sum_{\tau=m+1}^{n}\delta_{q,\tau-m}(v)\right] \\
& + in_m^{-1}\sum_{\tau=m+1}^{n}Z_{q,\tau,i}\delta_{q,\tau-m}(v) - i\left[n_m^{-1}\sum_{\tau=m+1}^{n}Z_{q,\tau,i}\right]\left[n_m^{-1}\sum_{\tau=m+1}^{n}\delta_{q,\tau-m}(v)\right] \\
& + in_m^{-1}\sum_{\tau=m+1}^{n}(Z_{\tau,i} - Z_{q,\tau,i})\psi_{q,\tau-m}(v) - i\left[n_m^{-1}\sum_{\tau=m+1}^{n}(Z_{\tau,i} - Z_{q,\tau,i})\right]\left[n_m^{-1}\sum_{\tau=m+1}^{n}\psi_{q,\tau-m}(v)\right] \\
= \ & i\left[\widehat{B}_{1mq}(v) - \widehat{B}_{2mq}(v) + \widehat{B}_{3mq}(v) - \widehat{B}_{4mq}(v) + \widehat{B}_{5mq}(v) - \widehat{B}_{6mq}(v)\right]
\end{aligned}
$$

Following the same reasoning as that of Theorem C.1 and noting that $E[Z_\tau \mid I_{\tau-1}] = 0$ *a.s.* and $E[Z_{q,\tau} \mid I_{\tau-1}] = 0$ *a.s.*, we have

$$
\begin{aligned}
p^{-\frac{1}{2}}\widehat{A}_{1,q} & \leqq 8p^{-\frac{1}{2}}\sum_{a'=1}^{6}\sum_{\tau=1}^{n-1}k^2(m/p)n_m\int\left|\widehat{B}_{a'mq}(v)\right|^2 dW(v) \\
& = O_p(p^{\frac{1}{2}}/q^{\kappa}) = o_p(1)
\end{aligned}
$$

given Assumption 4.4.2, $q/p \to \infty$, and $\kappa \geqq 1$. Further, by Cauchy-Schwartz inequality,

$$
\begin{aligned}
p^{-\frac{1}{2}}\widehat{A}_{2,q} & = 2p^{-\frac{1}{2}}\sum_{a'=1}^{6}\sum_{\tau=1}^{n-1}k^2(m/p)n_m Re\int\widehat{B}_{a'mq}(v)\widetilde{\sigma}^{(1,0)}_{q,m,i}(0,v)^* dW(v) \\
& = O_p(p^{\frac{1}{2}}/q^{\kappa}) = o_p(1)
\end{aligned}
$$

This completes the proof of Theorem C.2. Q.E.D.

Proof of Theorem C.3. I shall show Proposition C.3 and C.4 below. Q.E.D.

Proposition C.3: Let $\widetilde{\sigma}^{(1,0)}_{q,m}(0,v)$ be defined as $\widetilde{\sigma}^{(1,0)}_{m}(0,v)$, and let $\widetilde{C}_{0q}(p)$ be defined as $\widetilde{C}_0(p)$, with $\{Z_{q,\tau}\}_{\tau=1}^n$ replacing $\{Z_\tau\}_{\tau=1}^n$. Then under the conditions of Theorem 4.1.1,

$$\begin{aligned} & p^{-1/2}\sum_{m=1}^{n-1} k^2(m/p)n_m \int \left\|\widetilde{\sigma}^{(1,0)}_{q,m}(0,v)\right\|^2 dW(v) \\ = \;& p^{-1/2}\widetilde{C}_{0q}(p) + p^{-1/2}\widetilde{V}_q + o_p(1) \end{aligned} \tag{C.19}$$

where

$$\widetilde{V}_q = \sum_{a=i and (i,j) with i,j=1,\cdots d} \widetilde{V}_{q,a} and \widetilde{C}_{0q}(p) = \sum_{a=i and (i,j) with i,j=1,\cdots d} \widetilde{C}_{0q,a}(p)$$

and

$$\begin{aligned} \widetilde{V}_{q,a} &= \sum_{\tau=2q+2}^{n} Z_{q,\tau,a} \sum_{m=1}^{q} a_n(m) \int \psi_{q,\tau-m}(v) \left[\sum_{s=1}^{\tau-2q-1} Z_{q,s,a}\psi^*_{q,s-m}(v)\right] dW(v) \\ \widetilde{C}_{0q,a}(p) &= \sum_{m=1}^{n-1} k^2(m/p)\frac{1}{n-m}\sum_{\tau=m+1}^{n-1} Z^2_{q,\tau,a} \int |\psi_{\tau-m}(v)|^2 dW(v) \end{aligned}$$

Proposition C.4: Let $\widetilde{D}_{0q}(p)$ be defined as $\widetilde{D}_0(p)$ with $\{Z_{q,\tau}\}_{\tau=1}^n$ replacing $\{Z_\tau\}_{\tau=1}^n$. Then

$$\left[\widetilde{D}_{0q}(p)\right]^{-1/2}\widetilde{V}_q \to^d N(0,1)$$

Proof of Proposition C.3: Recall that $\widetilde{\sigma}^{(1,0)}_{q,m,a}(0,v) = n_m^{-1}\sum_{\tau=m+1}^n Z_{q,\tau,a}\psi_{q,\tau-m}(v)$, where $\psi_{q,\tau}(v) \equiv e^{iv'Z_{q,\tau}} - \varphi_q(v)$ and $\varphi_q(v) = E\left(e^{iv'Z_{q,\tau}}\right)$. Then

$$\begin{aligned} & \sum_{m=1}^{n-1} k^2(m/p)n_m \int \left\|\widetilde{\sigma}^{(1,0)}_{q,m}(0,v)\right\|^2 dW(v) \\ = \;& \sum_{m=1}^{n-1} k^2(m/p)n_m \int \sum_{a=i and (i,j) with i,j=1,\cdots d} \left|\widetilde{\sigma}^{(1,0)}_{q,m,a}(0,v)\right|^2 dW(v) \\ = \;& \sum_{a=i and (i,j) with i,j=1,\cdots d} \left[\sum_{m=1}^{n-1} k^2(m/p)n_m \int \left|\widetilde{\sigma}^{(1,0)}_{q,m,a}(0,v)\right|^2 dW(v)\right] \end{aligned}$$

Henceforth, to prove (C.19), it is sufficient to show that

$$
\begin{aligned}
& p^{-1/2} \sum_{m=1}^{n-1} k^2(m/p) n_m \int \left| \widetilde{\sigma}^{(1,0)}_{q,m,a}(0,v) \right|^2 dW(v) \\
= \; & p^{-1/2} \widetilde{C}_{0q,a}(p) + p^{-1/2} \widetilde{V}_{q,a} + o_p(1) \qquad \text{(C.20)}
\end{aligned}
$$

To show (C.20), I first decompose

$$
\begin{aligned}
& \sum_{m=1}^{n-1} k^2(m/p) n_m \int \left| \widetilde{\sigma}^{(1,0)}_{q,m,a}(0,v) \right|^2 dW(v) \\
= \; & \sum_{m=1}^{n-1} a_n(m) \int \left| \sum_{\tau=1}^{n} Z_{q,\tau,a} \psi_{q,\tau-m}(v) \right|^2 dW(v) \\
& + \sum_{m=1}^{n-1} a_n(m) \int \left| \sum_{\tau=1}^{m} Z_{q,\tau,a} \psi_{q,\tau-m}(v) \right|^2 dW(v) \\
& -2Re \sum_{m=1}^{n-1} a_n(m) \int \left[\sum_{\tau=1}^{n} Z_{q,\tau,a} \psi_{q,\tau-m}(v) \right] \left[\sum_{\tau=1}^{m} Z_{q,\tau,a} \psi_{q,\tau-m}(v) \right]^* dW(v \\
\equiv \; & \widetilde{Q}_q + \widetilde{R}_{1q} - 2Re\left(\widetilde{R}_{2q} \right) \qquad \text{(C.21)}
\end{aligned}
$$

Next write

$$
\begin{aligned}
\widetilde{Q}_q \; = \; & \sum_{m=1}^{n-1} a_n(m) \int \sum_{\tau=1}^{n} Z^2_{q,\tau,a} \left| \psi_{q,\tau-m}(v) \right|^2 dW(v) \\
& +2Re \sum_{m=1}^{n-1} a_n(m) \int \sum_{\tau=2}^{n} \sum_{s=1}^{\tau-1} Z_{q,\tau,a} Z_{q,s,a} \psi_{q,\tau-m}(v) \psi^*_{q,s-m}(v) dW(v) \\
\equiv \; & \widetilde{C}_q(p) + 2Re\left(\widetilde{U}_q \right) \qquad \text{(C.22)}
\end{aligned}
$$

and further decompose

$$
\widetilde{U}_q \; = \; \sum_{\tau=2q+2}^{n} Z_{q,\tau,a} \int \sum_{m=1}^{n-2} a_n(m) \psi_{q,\tau-m}(v) \sum_{s=1}^{\tau-2q-1} Z_{q,s,a} \psi^*_{q,s-m}(v) dW(v)
$$

$$+\sum_{\tau=2}^{n} Z_{q,\tau,a} \int \sum_{m=1}^{n-2} a_n(m)\psi_{q,\tau-m}(v) \sum_{s=\max(1,\tau-2q)}^{\tau-1} Z_{q,s,a}\psi^*_{q,s-m}(v)dW(v)$$

$$\equiv \widetilde{U}_{1q} + \widetilde{R}_{3q} \tag{C.23}$$

where in the first term $\widetilde{U}_{1q}$, we have $\tau - s > 2q$ so that $\{Z_{q,\tau,a}, \psi_{q,\tau-m}(v)\}_{m=1}^{q}$ is independent of $\{Z_{q,s,a}, \psi_{q,s-m}(v)\}_{m=1}^{q}$ for q sufficiently large. In the second term $\widetilde{R}_{3q}$, we have $0 < \tau - s \leq 2q$. Finally, write

$$\widetilde{U}_{1q} = \sum_{\tau=2q+2}^{n} Z_{q,\tau,a} \sum_{m=1}^{q} a_n(m) \int \psi_{q,\tau-m}(v) \sum_{s=1}^{\tau-2q-1} Z_{q,s,a}\psi^*_{q,s-m}(v)dW(v)$$

$$+ \sum_{\tau=2q+2}^{n} Z_{q,\tau,a} \sum_{m=q+1}^{n-1} a_n(m) \int \psi_{q,\tau-m}(v) \sum_{s=1}^{\tau-2q-1} Z_{q,s,a}\psi^*_{q,s-m}(v)dW(v)$$

$$\equiv \widetilde{V}_{q,a} + \widetilde{R}_{4q} \tag{C.24}$$

where the first term $\widetilde{V}_{q,a}$ is contributed by the lad orders m from 1 to q; and the second term $\widetilde{R}_{4q}$ is from lag orders $m > q$. It follows from (C.21) to (C.24) that

$$\sum_{m=1}^{n-1} k^2(m/p)n_m \int \left|\widetilde{\sigma}^{(1,0)}_{q,m,a}(0,v)\right|^2 dW(v)$$

$$= \widetilde{C}_q(p) + 2Re\left(\widetilde{V}_{q,a}\right) + \widetilde{R}_{1q} - 2Re\left(\widetilde{R}_{2q} - \widetilde{R}_{3q} - \widetilde{R}_{4q}\right)$$

It suffices to show Lemmas C.7-C.11 below, which imply $p^{-1/2}\left[\widetilde{C}_q(p) - \widetilde{C}_{0q,a}(p)\right] = o_p(1)$ and $p^{-1/2}\widetilde{R}_{a'q} = o_p(1)$ for $a' = 1,2,3,4$ given $q = p^{1+\frac{1}{4b-2}}\left(\ln^2 n\right)^{\frac{1}{2b-1}}$ and $p = cn^{\lambda}$ for $0 < \lambda < \left(3 + \frac{1}{4b-2}\right)^{-1}$. Q.E.D.

Lemma C.7: Let $\widetilde{C}_q(p)$ be defined as in (C.22). Then $\widetilde{C}_q(p) - \widetilde{C}_{0q,a}(p) = O_p(p^2/n)$.

Lemma C.8: Let $\widetilde{R}_{1q}$ be defined as in (C.21). Then $\widetilde{R}_{1q} = O_p(p^2/n)$.

Lemma C.9: Let $\widetilde{R}_{2q}$ be defined as in (C.21). Then $\widetilde{R}_{2q} = O_p(p^{\frac{3}{2}}/n^{\frac{1}{2}})$.

Lemma C.10: Let $\widetilde{R}_{3q}$ be defined as in (C.23). Then $\widetilde{R}_{3q} = O_p(q^{\frac{1}{2}}p/n^{\frac{1}{2}})$.

Lemma C.11: Let $\widetilde{R}_{4q}$ be defined as in (C.24). Then $\widetilde{R}_{4q} = O_p(p^{2b}\ln(n)/q^{2b-1})$.

Proof of Lemma C.7: By Markov's inequality and $E\left|\widetilde{C}_q(p) - \widetilde{C}_{0q,a}(p)\right| \leq Cp^2/n$ given $\sum_{m=1}^{n-1}(m/p)\,a_n(m) = O(p/n)$. Q.E.D.

Proof of Lemma A.8: By the *m.d.s.* property of $\{Z_{q,\tau}, \mathfrak{F}_{\tau-1}\}$ where $\mathfrak{F}_{\tau-1}$ is the sigma-field generated by $\{Z_{\tau-m}\}_{m=1}^{\infty}$, we can obtain

$$\begin{aligned} & E\int\left|\sum_{\tau=1}^{m} Z_{q,\tau,a}\psi_{q,\tau-m}(v)\right|^2 dW(v) \\ = & \sum_{\tau=1}^{m}\int E\left[Z_{q,\tau,a}^2\left|\psi_{q,\tau-m}(v)\right|^2\right]dW(v) \leq Cm \end{aligned}$$

The result then follows from Markov's inequality and $\sum_{m=1}^{n-1}(m/p)\,a_n(m) = O(p/n)$ given Assumption 4.4.6. Q.E.D.

Proof of Lemma C.9: The proof is similar to that of Lemma C.8, with the fact that

$$E\left|\int\left[\sum_{\tau=1}^{n} Z_{q,\tau,a}\psi_{q,\tau-m}(v)\right]\left[\sum_{\tau=1}^{m} Z_{q,\tau,a}\psi_{q,\tau-m}(v)\right]^* dW(v)\right| \leq C\,(mn)^{1/2}$$

given Assumption 4.4.6. Q.E.D.

Proof of Lemma C.10: By the *m.d.s.* property of $\{Z_{q,\tau}, \mathfrak{F}_{\tau-1}\}$, Minkowski's inequality and (C.11), we have

$$\begin{aligned} & E\left|\widetilde{R}_{3q}\right|^2 \\ = & \sum_{\tau=2}^{n} E\left|\sum_{m=1}^{n-1} a_n(m)\int Z_{q,\tau,a}\psi_{q,\tau-m}(v)\sum_{s=\max(1,\tau-2q)}^{\tau-1} Z_{q,s,a}\psi_{q,s-m}^*(v)dW(v)\right|^2 \\ \leq & \sum_{\tau=2}^{n}\left[\sum_{m=1}^{n-1} a_n(m)\int\left(E\left|Z_{q,\tau,a}\psi_{q,\tau-m}(v)\sum_{s=\max(1,\tau-2q)}^{\tau-1} Z_{q,s,a}\psi_{q,s-m}^*(v)\right|^2\right)^{\frac{1}{2}} dW(v)\right]^2 \end{aligned}$$

$$\leq \; 2Cnq\left[\sum_{m=1}^{n-1} a_n(m)\right]^2 = O(qp^2/n)$$

This finishes the proof of Lemma C.10. Q.E.D.

Proof of Lemma C.11: By the *m.d.s.* property of $\{Z_{q,\tau}, \mathfrak{F}_{\tau-1}\}$ and Minkowski's inequality,

$$\begin{aligned}
& E\left|\widetilde{R}_{4q}\right|^2 \\
= & \sum_{\tau=2q+2}^{n} E\left|\sum_{m=q+1}^{n-1} a_n(m)\int Z_{q,\tau,a}\psi_{q,\tau-m}(v)\sum_{s=1}^{\tau-2q-1} Z_{q,s,a}\psi^*_{q,s-m}(v)dW(v)\right|^2 \\
\leq & \sum_{\tau=2q+2}^{n}\left[\sum_{m=q+1}^{n-1} a_n(m)\int\left(E\left|Z_{q,\tau,a}\psi_{q,\tau-m}(v)\sum_{s=1}^{\tau-2q-1} Z_{q,s,a}\psi^*_{q,s-m}(v)\right|^2\right)^{\frac{1}{2}} dW(v)\right]^2 \\
\leq & \; Cn^2\left[\sum_{m=q+1}^{n-1} a_n(m)\right]^2 \\
\leq & \; Cn^2\left[\sum_{m=q+1}^{n-1} (m/p)^{-2b}\, n_m^{-1}\right]^2 = O(p^{4b}\ln^2(n)/q^{4b-2})
\end{aligned}$$

given that fact that $k(z) \leq C\,|z|^{-b}$ as $z \to \infty$ from Assumption 4.4.6. Q.E.D.

Proof of Proposition C.4: From (C.19), $\widetilde{V}_q = \sum_{a=iand(i,j)withi,j=1,\cdots d} \widetilde{V}_{q,a}$. We rewrite $\widetilde{V}_q = \sum_{\tau=2q+2}^{n} V_q(\tau)$, where

$$\begin{aligned}
V_q(\tau) &= \sum_{a=iand(i,j)withi,j=1,\cdots d} V_{q,a}, V_{q,a} \\
&= Z_{q,\tau,a}\sum_{m=1}^{q} a_n(m)\int \psi_{q,\tau-m}(v)H_{m,.\tau-2q-1,a}(v)dW(v)
\end{aligned}$$

and

$$H_{m,.\tau-2q-1,a}(v) = \sum_{s=1}^{\tau-2q-1} Z_{q,s,a}\psi^*_{q,s-m}(v)$$

Then I will apply the martingale central limit theorem (Brown, 1971), which states that $var\left(2Re\widetilde{V}_q\right)^{-\frac{1}{2}} 2Re\widetilde{V}_q \to^d N(0,1)$ if

$$var\left(2Re\widetilde{V}_q\right)^{-1} \sum_{\tau=1}^{n} \left[2ReV_q(\tau)\right]^2 1\left[\left|2ReV_q(\tau)\right| > \eta \cdot var\left(2Re\widetilde{V}_q\right)^{\frac{1}{2}}\right] \to 0, \qquad \text{(C.25)}$$

for any $\eta > 0$, and

$$var\left(2Re\widetilde{V}_q\right)^{-1} \sum_{\tau=1}^{n} E\left[2ReV_q^2(\tau)|\mathfrak{F}_{\tau-1}\right] \to^p 1 \qquad \text{(C.26)}$$

First, let's compute $var\left(2Re\widetilde{V}_q\right)^{-1}$. By the *m.d.s.* property of $\{Z_{q,\tau}, \mathfrak{F}_{\tau-1}\}$ under H_0 and independence between $Z_{q,\tau}$ and $\{Z_{\tau-m-1}\}_{m=q}^{\infty}$ for q sufficiently large, we have

$$\begin{aligned}
& E\left(\widetilde{V}_q^2\right) \\
= & \sum_{\tau=2q+2}^{n} \sum \sum_{a,a'=i,(i,j),i,j=1,\cdots d} E\Bigg[Z_{q,\tau,a}Z_{q,\tau,a'} \sum_{m=1}^{q}\sum_{l=1}^{q} a_n(m)a_n(l) \\
& \int\int \psi_{q,\tau-m}(v)\psi_{q,\tau-l}(u)H_{m,.\tau-2q-1,a}(v)H_{m,.\tau-2q-1,a'}(u)dW(v)dW(u)\Bigg] \\
= & \sum \sum_{a,a'=i,(i,j);i,j=1,\cdots d} \sum_{m=1}^{q}\sum_{l=1}^{q} a_n(m)a_n(l) \int\int \sum_{\tau=2q+2}^{n} \sum_{s=1}^{\tau-2q-1} \\
& \times E\left[Z_{q,\tau,a}Z_{q,\tau,a'}\psi_{q,\tau-m}(v)\psi_{q,\tau-l}(u)\right] \\
& \times E\left[Z_{q,s,a}Z_{q,s,a'}\psi^*_{q,s-m}(v)\psi^*_{q,s-l}(u)\right] dW(v)dW(u) \\
= & \frac{1}{2}\sum \sum_{a,a'=i,(i,j);i,j=1,\cdots d} \sum_{m=1}^{q}\sum_{l=1}^{q} k^2(m/p)k^2(l/p) \\
& \times \int\int \left|E\left[Z_{q,0,a}Z_{q,0,a'}\psi_{q,-m}(v)\psi_{q,-l}(u)\right]\right|^2 dW(v)dW(u)\,[1+o(1)]
\end{aligned}$$

Similarly, we can obtain

$$E\left(\widetilde{V}_q^*\right)^2$$

$$= \frac{1}{2}\sum \sum_{a,a'=i,(i,j);i,j=1,\cdots d} \sum_{m=1}^{q}\sum_{l=1}^{q} k^2(m/p)k^2(l/p).$$
$$\times \int\int \left|E\left[Z_{q,0,a}Z_{q,0,a'}\psi_{q,-m}(v)\psi_{q,-l}(u)\right]\right|^2 dW(v)dW(u)\left[1+o(1)\right]$$

and

$$E\left|\widetilde{V}_q\right|^2$$
$$= \frac{1}{2}\sum \sum_{a,a'=i,(i,j);i,j=1,\cdots d} \sum_{m=1}^{q}\sum_{l=1}^{q} k^2(m/p)k^2(l/p)$$
$$\times \int\int \left|E\left[Z_{q,0,a}Z_{q,0,a'}\psi_{q,-m}(v)\psi_{q,-l}(u)\right]\right|^2 dW(v)dW(u)\left[1+o(1)\right]$$

Because $W(\cdot)$ wights sets symmetric about zero equally, we have $E\left|\widetilde{V}_q\right|^2 = E\left(\widetilde{V}_q^2\right) = E\left(\widetilde{V}_q^*\right)^2$. Hence

$$var\left(2Re\widetilde{V}_q\right)$$
$$= 2E\left|\widetilde{V}_q\right|^2 + E\left(\widetilde{V}_q^2\right) + E\left(\widetilde{V}_q^*\right)^2$$
$$= 2\sum \sum_{a,a'=i,(i,j);i,j=1,\cdots d} \sum_{m=1}^{q}\sum_{l=1}^{q} k^2(m/p)k^2(l/p)$$
$$\times \int\int \left|E\left[Z_{0,a}Z_{0,a'}\psi_{-m}(v)\psi_{-l}(u)\right]\right|^2 dW(v)dW(u)\left[1+o(1)\right] \qquad \text{(C.27)}$$

where we have used that fact that $E\left[Z_{q,0,a}Z_{q,0,a'}\psi_{q,-m}(v)\psi_{q,-l}(u)\right] \to E\left[Z_{0,a}Z_{0,a'}\psi_{-m}(v)\psi_{-l}(u)\right]$ as $q\to\infty$ given Assumption 4.4.2. Put $C(0,m,l) \equiv E\left[\left(Z_{0,a}Z_{0,a'} - \sigma(a,a')\right)\psi_{-m}(v)\psi_{-l}(u)\right]$. Then

$$E\left[Z_{0,a}Z_{0,a;}\psi_{-m}(v)\psi_{-l}(u)\right]$$
$$= C(0,m,l) + \sigma(a,a')\sigma_{l-m}(v,u)\left|E\left[Z_{0,a}Z_{0,a'}\psi_{-m}(v)\psi_{-l}(u)\right]\right|^2$$
$$= |C(0,m,l)|^2 + \sigma(a,a')^2\left|\sigma_{l-m}(v,u)\right|^2 + 2\sigma(a,a')Re\left[C(0,m,l)\sigma_{l-m}^*(v,u)\right]$$

Given $\sum_{m=-\infty}^{\infty}\sum_{l=-\infty}^{\infty}|C(0,m,l)| \leq C$ and $|k(\cdot)| \leq 1$, we have

$$
\begin{aligned}
& var\left(2Re\widetilde{V}_q\right) \\
= & \sum \sum_{a,a'=i,(i,j);i,j=1,\cdots d} 2\sigma(a,a')^2 \sum_{m=1}^{q}\sum_{l=1}^{q} k^2(m/p)k^2(l/p) \\
& \times \int\int |\sigma_{l-m}(v,u)|^2\, dW(v)dW(u)\,[1+o(1)] \\
= & \sum \sum_{a,a'=i,(i,j);i,j=1,\cdots d} 2\sigma(a,a')^2 p \sum_{c=1-q}^{q-1}\left[p^{-1}\sum_{m=c+1}^{q} k^2(m/p)k^2\left[(m-c)/p\right]\right] \\
& \times \int\int |\sigma_{c}(v,u)|^2\, dW(v)dW(u)\,[1+o(1)] \\
= & \sum \sum_{a,a'=i,(i,j);i,j=1,\cdots d} 2\sigma(a,a')^2 p \int_0^{\infty} k^4(z)dz \sum_{c=-\infty}^{\infty} \\
& \times \int\int |\sigma_{c}(v,u)|^2\, dW(v)dW(u)\,[1+o(1)] \\
= & \sum \sum_{a,a'=i,(i,j);i,j=1,\cdots d} 4\pi\sigma(a,a')^2 p \int_0^{\infty} k^4(z)dz \\
& \times \int\int\int_{-\pi}^{\pi} |f(\omega,v,u)|^2\, d\omega dW(v)dW(u)\,[1+o(1)]
\end{aligned}
$$

where we used the fact that for many given c, $p^{-1}\sum_{m=c+1}^{q} k^2(m/p)k^2\left[(m-c)/p\right] \to \int_0^{\infty} k^4(z)dz$ as $p\to\infty$ and $q/p\to 0$.

I now verify condition (C.25). By $E\left|H_{m,.\tau-2q-1,a}(v)\right|^4 \le C\tau^2$ for $1\le m\le q$ given the *m.d.s.* property of $\{Z_{q,\tau},\mathfrak{F}_{\tau-1}\}$ and Rosenthal's inequality for martingale(see Hall and Heyde, 1980, p.23),

$$
\begin{aligned}
& E\left|V_{q,a}(\tau)\right|^4 \\
\le & \left[\sum_{m=1}^{q} a_n(m)\int\left(E\left|Z_{q,\tau,a}\psi_{q,\tau-m}(v)H_{m,.\tau-2q-1,a}(v)\right|^4\right)^{\frac{1}{4}} dW(v)\right]^4 \\
\le & C\tau^2\left[\sum_{m=1}^{q} a_n(m)\right]^4 = O(p^4\tau^2/n^4) \qquad \text{(C.28)}
\end{aligned}
$$

Then recall that $V_q(\tau) = \sum_{a=iand(i,j)withi,j=1,\cdots d} V_{q,a}$ and use Jensen's inequality, we have

$$\begin{aligned}
E\left|V_q(\tau)\right|^4 &= E\left|\sum_{a=iand(i,j),i,j=1,\cdots d} V_{q,a}\right|^4 \\
&\leq E\left[\sum_{a=iand(i,j)withi,j=1,\cdots d} \left|V_{q,a}(\tau)\right|\right]^4 \\
&= E\left\{(d')^4\left[\sum_{a=iand(i,j),i,j=1,\cdots d} \left|V_{q,a}(\tau)\right| \frac{1}{d'}\right]^4\right\} \\
&\leq E\left\{(d')^4 \sum_{a=iand(i,j),i,j=1,\cdots d} \left|V_{q,a}(\tau)\right|^4 \frac{1}{d'}\right\} \\
&= (d')^3 \sum_{a=iand(i,j),i,j=1,\cdots d} E\left|V_{q,a}(\tau)\right|^4 = O(p^4\tau^2/n^4)
\end{aligned}$$

where the last equality uses (C.28). It then follows that

$$\sum_{\tau=2q+2}^{n} E\left|V_q(\tau)\right|^4 = O(p^4/n) = o(p^2) given p^2/n \to 0.$$

Therefore (C.25) is proved.

Next I turn to verify condition (C.26). Let $\sigma^2_{q,\tau}(a,a') \equiv E\left(Z_{q,\tau,a}Z_{q,\tau,a'}|\mathfrak{F}_{\tau-1}\right)$. Then

$$\begin{aligned}
& E\left[V_q^2(\tau)|\mathfrak{F}_{\tau-1}\right] \\
= & \sum \sum_{a,a'=i,(i,j);i,j=1,\cdots d} \sum_{m=1}^{q}\sum_{l=1}^{q} a_n(m)a_n(l) \int\int \sigma^2_{q,\tau}(a,a')\psi_{q,\tau-m}(v)\psi_{q,\tau-l}(u) \\
& \times H_{m,.\tau-2q-1,a}(v)H_{l,.\tau-2q-1,a'}(u)dW(v)dW(u) \\
= & \sum \sum_{a,a'=i,(i,j);i,j=1,\cdots d} \sum_{m=1}^{q}\sum_{l=1}^{q} a_n(m)a_n(l) \int\int E\left[\sigma^2_{q,\tau}(a,a')\psi_{q,\tau-m}(v)\psi_{q,\tau-l}(u)\right] \\
& \times H_{m,.\tau-2q-1,a}(v)H_{l,.\tau-2q-1,a'}(u)dW(v)dW(u) \\
& + \sum \sum_{a,a'=i,(i,j);i,j=1,\cdots d} \sum_{m=1}^{q}\sum_{l=1}^{q} a_n(m)a_n(l) \int\int \widetilde{Z}^{m,l}_{q,\tau,aa'}(v,u)H_{m,.\tau-2q-1,a}(v) \\
& \times H_{l,.\tau-2q-1,a'}(u)dW(v)dW(u) \\
\equiv & S_{1q}(\tau) + V_{1q}(\tau) \qquad (C.29)
\end{aligned}$$

where

$$\widetilde{Z}^{m,l}_{q,\tau,aa'}(v,u) \equiv \sigma^2_{q,\tau}(a,a')\psi_{q,\tau-m}(v)\psi_{q,\tau-l}(u) - E\left[\sigma^2_{q,\tau}(a,a')\psi_{q,\tau-m}(v)\psi_{q,\tau-l}(u)\right]$$

We further decompose

$$\begin{aligned}
& S_{1q}(\tau) \\
= & \sum \sum_{a,a'=i,(i,j);i,j=1,\cdots d} \sum_{m=1}^{q}\sum_{l=1}^{q} a_n(m)a_n(l) \int\int E\left[\sigma^2_{q,\tau}(a,a')\psi_{q,\tau-m}(v)\psi_{q,\tau-l}(u)\right] \\
& \times E\left[H_{m,.\tau-2q-1,a}(v)H_{l,.\tau-2q-1,a'}(u)\right] dW(v)dW(u) \\
& + \sum \sum_{a,a'=i,(i,j);i,j=1,\cdots d} \sum_{m=1}^{q}\sum_{l=1}^{q} a_n(m)a_n(l) \int\int E\left[\sigma^2_{q,\tau}(a,a')\psi_{q,\tau-m}(v)\psi_{q,\tau-l}(u)\right] \\
& \times \left\{H_{m,.\tau-2q-1,a}(v)H_{l,.\tau-2q-1,a'}(u) - E\left[H_{m,.\tau-2q-1,a}(v)H_{l,.\tau-2q-1,a'}(u)\right]\right\} dW(v)dW(u) \\
\equiv & E\left[V_q^2(\tau)\right] + S_{2q}(\tau) \qquad \text{(C.30)}
\end{aligned}$$

where

$$\begin{aligned}
& E\left[V_q^2(\tau)\right] \\
\equiv & \sum \sum_{a,a'=i,(i,j);i,j=1,\cdots d} \sum_{m=1}^{q}\sum_{l=1}^{q} (\tau-q-1)a_n(m)a_n(l) \\
& \times \int\int \left| E\left[\sigma^2_{q,\tau}(a,a')\psi_{q,\tau-m}(v)\psi_{q,\tau-l}(u)\right]\right| dW(v)dW(u)
\end{aligned}$$

Put

$$Z^{m,l}_{q,s,aa'}(v,u) \equiv Z_{q,s,a}Z_{q,s,a'}\psi_{q,s-m}(v)\psi_{q,s-l}(u) - E\left[Z_{q,s,a}Z_{q,s,a'}\psi_{q,s-m}(v)\psi_{q,s-l}(u)\right]$$

Then write

$$\begin{aligned}
& S_{2q}(\tau) \\
= & \sum \sum_{a,a'=i,(i,j);i,j=1,\cdots d} \sum_{m=1}^{q}\sum_{l=1}^{q} a_n(m)a_n(l) \int\int E\left[Z_{q,\tau,a}Z_{q,\tau,a'}\psi_{q,\tau-m}(v)\psi_{q,\tau-l}(u)\right] \\
& \times \sum_{s=1}^{\tau-2q-1} Z^{m,l}_{q,s,aa'}(v,u)dW(v)dW(u)
\end{aligned}$$

$$
\begin{aligned}
& +\sum_{a,a'=i,(i,j);i,j=1,\cdots d}\sum_{m=1}^{q}\sum_{l=1}^{q} a_n(m)a_n(l)\int\int E\left[Z_{q,\tau,a}Z_{q,\tau,a'}\psi_{q,\tau-m}(v)\psi_{q,\tau-l}(u)\right] \\
& \times \sum_{s=2}^{\tau-2q-1}\sum_{c=1}^{s-1} Z_{q,s,a}\psi_{q,s-m}(v)Z_{q,c,a'}\psi_{q,c-l}(u)dW(v)dW(u) \\
\equiv\ & V_{2q}(\tau)+S_{3q}(\tau) \qquad \text{(C.31)}
\end{aligned}
$$

where

$$
\begin{aligned}
& S_{3q}(\tau) \\
=\ & \sum_{a,a'=i,(i,j);i,j=1,\cdots d}\sum_{m=1}^{q}\sum_{l=1}^{q} a_n(m)a_n(l)\int E\left[Z_{q,\tau,a}Z_{q,\tau,a'}\psi_{q,\tau-m}(v)\psi_{q,\tau-l}(u)\right] \\
& \times \sum_{s=2}^{\tau-2q-1}\sum_{0<s-c\leq 2q} Z_{q,s,a}\psi_{q,s-m}(v)Z_{q,c,a'}\psi_{q,c-l}(u)dW(v)dW(u) \\
& +\sum_{a,a'=i,(i,j);i,j=1,\cdots d}\sum_{m=1}^{q}\sum_{l=1}^{q} a_n(m)a_n(l)\int E\left[Z_{q,\tau,a}Z_{q,\tau,a'}\psi_{q,\tau-m}(v)\psi_{q,\tau-l}(u)\right] \\
& \times \sum_{s=2}^{\tau-2q-1}\sum_{s-c>2q} Z_{q,s,a}\psi_{q,s-m}(v)Z_{q,c,a'}\psi_{q,c-l}(u)dW(v)dW(u) \\
\equiv\ & V_{3q}(\tau)+V_{4q}(\tau) \qquad \text{(C.32)}
\end{aligned}
$$

It follows from (C.29)-(C.32) that

$$
\sum_{\tau=2q+2}^{n}\left\{E\left[V_q^2(\tau)|\mathfrak{F}_{\tau-1}\right]-E\left[V_q^2(\tau)\right]\right\}=\sum_{h=1}^{4}\sum_{\tau=2q+2}^{n}V_{hq}(\tau)
$$

Then it is sufficient to show Lemmas C.12-C.15 below, which imply that

$$
E\left|\sum_{\tau=2q+2}^{n}\left\{E\left[V_q^2(\tau)|\mathfrak{F}_{\tau-1}\right]-E\left[V_q^2(\tau)\right]\right\}\right|^2=o(p^2)
$$

given $q=p^{1+\frac{1}{4b-2}}\left(\ln^2 n\right)^{\frac{1}{2b-1}}$ and $p=cn^{\lambda}$ for $0<\lambda<\left(3+\frac{1}{4b-2}\right)^{-1}$. Thus condition (C.26) holds and so $M_{0q}(p)\to^{d}N(0,1)$ by Brown's theorem. Q.E.D.

Lemma C.12: Let $V_{1q}(\tau)$ be defined as in (C.29). Then $E\left|\sum_{\tau=2q+2}^{n}V_{1q}(\tau)\right|^2=O(qp^4/n)$.

Lemma C.13: Let $V_{2q}(\tau)$ be defined as in (C.31). Then $E\left|\sum_{\tau=2q+2}^{n} V_{2q}(\tau)\right|^2 = O(qp^4/n)$.

Lemma C.14: Let $V_{3q}(\tau)$ be defined as in (C.32). Then $E\left|\sum_{\tau=2q+2}^{n} V_{3q}(\tau)\right|^2 = O(qp^4/n)$.

Lemma C.15: Let $V_{4q}(\tau)$ be defined as in (C.32). Then $E\left|\sum_{\tau=2q+2}^{n} V_{4q}(\tau)\right|^2 = O(p)$.

Proof of Lemma C.12: Let

$$\begin{aligned} V_{1q,aa'}(\tau) \equiv & \sum_{m=1}^{q}\sum_{l=1}^{q} a_n(m)a_n(l)\int\int \widetilde{Z}_{q,\tau,aa'}^{m,l}(v,u) \\ & \times H_{m,.\tau-2q-1,a}(v)H_{l,.\tau-2q-1,a'}(u)dW(v)dW(u) \end{aligned}$$

Then from (C.29),

$$V_{1q}(\tau) = \sum \sum_{a,a'=i,(i,j);i,j=1,\cdots d} V_{1q,aa'}(\tau)$$

Recall the definition of $\widetilde{Z}_{q,\tau,aa'}^{m,l}(v,u)$ in (C.29). Noting that $\widetilde{Z}_{q,\tau,aa'}^{m,l}(v,u)$ is independent of $\left\{H_{m,.\tau-2q-1,a}(v)H_{l,.\tau-2q-1,a'}(u)\right\}$ and that $\widetilde{Z}_{q,\tau,aa'}^{m,l}(v,u)$ is independent of $\widetilde{Z}_{q,c,aa'}^{m,l}(v,u)$ for $\tau - c > 2q$ and $1 \le m,l \le q$, we have

$$\begin{aligned} & E\left|\sum_{\tau=2q+2}^{n} \widetilde{Z}_{q,\tau,aa'}^{m,l}(v,u)H_{m,.\tau-2q-1,a}(v)H_{l,.\tau-2q-1,a'}(u)\right|^2 \\ \le & \sum_{\tau=2q+2}^{n}\sum_{|\tau-c|\le 2q} E\left|\widetilde{Z}_{q,\tau,aa'}^{m,l}(v,u)\widetilde{Z}_{q,c,aa'}^{m,l}(v,u)\right| \\ & \times\left(E\left|H_{m,.\tau-2q-1,a}(v)\right|^4\right)^{\frac{1}{4}}\left(E\left|H_{l,.\tau-2q-1,a'}(u)\right|^4\right)^{\frac{1}{4}} \\ & \times\left(E\left|H_{m,.c-2q-1,a}(v)\right|^4\right)^{\frac{1}{4}}\left(E\left|H_{l,.c-2q-1,a'}(u)\right|^4\right)^{\frac{1}{4}} \\ = & O(n^3q) \end{aligned}$$

where we have used the fact that $E\left|H_{m,.\tau-2q-1,a}(v)\right|^4 \leq Cn^2$ for $1 \leq m \leq q$ and any a. It then follows by Minkowski's inequality and (C.11) that

$$
\begin{aligned}
& E\left|\sum_{\tau=2q+2}^{n} V_{1q,aa'}(\tau)\right|^2 \\
\leq & \left[\sum_{m=1}^{q}\sum_{l=1}^{q} a_n(m)a_n(l)\right. \\
& \left.\left(E\left|\sum_{\tau=2q+2}^{n}\int\int \widetilde{Z}^{m,l}_{q,\tau,aa'}(v,u)H_{m,.\tau-2q-1,a}(v)H_{l,.\tau-2q-1,a'}(u)\right|^2\right)^{\frac{1}{2}}\right]^2 \\
= & O(qp^4/n) \qquad \text{(C.33)}
\end{aligned}
$$

An application of Jensen's inequality implies that

$$
\begin{aligned}
& E\left|\sum_{\tau=2q+2}^{n} V_{1q}(\tau)\right|^2 \\
= & E\left|\sum_{a,a'=i,(i,j);i,j=1,\cdots d}\left[\sum_{\tau=2q+2}^{n} V_{1q,aa'}(\tau)\right]\right|^2 \\
\leq & d' \sum_{a,a'=i,(i,j);i,j=1,\cdots d} E\left|\sum_{\tau=2q+2}^{n} V_{1q,aa'}(\tau)\right|^2 = O(qp^4/n) \qquad \text{(C.34)}
\end{aligned}
$$

where (C.33) is used in the last equality. This completes the proof of Lemma C.12. Q.E.D.

Proof of Lemma C.13: Let

$$
\begin{aligned}
V_{2q,aa'}(\tau) \equiv & \sum_{m=1}^{q}\sum_{l=1}^{q} a_n(m)a_n(l)\int\int E\left[Z_{q,\tau,a}Z_{q,\tau,a'}\psi_{q,\tau-m}(v)\psi_{q,\tau-l}(u)\right] \\
& \times \sum_{s=1}^{\tau-2q-1} Z^{m,l}_{q,s,aa'}(v,u)dW(v)dW(u)
\end{aligned}
$$

Then from (C.31), $V_{2q}(\tau) = a, a' = iand(i,j), i, j = 1, \cdots d \sum \sum V_{2q,aa'}(\tau)$. Recall the definition of $Z^{m,l}_{q,s,aa'}(v,u)$ in (C.31). Noting that $\left\{Z^{m,l}_{q,s,aa'}(v,u)\right\}^{q}_{m,l=1}$ is independent of $\left\{Z^{m,l}_{q,c,aa'}(v,u)\right\}^{q}_{m,l=1}$ for $|s-c| > 2q$ where q is sufficiently large, we have

$$E\left|\sum_{s=1}^{\tau-q-1} Z^{m,l}_{q,s,aa'}(v,u)\right|^2 = \sum_{s=1}^{\tau-q-1}\sum_{|s-c|\le 2q} E\left[Z^{m,l}_{q,s,aa'}(v,u)Z^{m,l}_{q,c,aa'}(v,u)\right] \le 2C\tau q \qquad \text{(C.35)}$$

It then follows that

$$\begin{aligned}
& E\left|\sum_{\tau=2q+2}^{n} V_{2q,aa'}(\tau)\right|^2 \\
\le\ & \left\{\sum_{\tau=2q+2}^{n}\left[E\left|V_{2q,aa'}(\tau)\right|^2\right]^{\frac{1}{2}}\right\}^2 \\
\le\ & \left\{\sum_{\tau=2q+2}^{n}\sum_{m=1}^{q}\sum_{l=1}^{q} a_n(m)a_n(l)\int\left|E\left[Z_{q,\tau,a}Z_{q,\tau,a'}\psi_{q,\tau-m}(v)\psi_{q,\tau-l}(u)\right]\right|\right. \\
& \left.\times\left(E\left|\sum_{s=1}^{\tau-2q-1} Z^{m,l}_{q,s,aa'}(v,u)\right|^2\right)^{\frac{1}{2}} dW(v)dW(u)\right\}^2 \\
=\ & O(qp^4/n) \qquad \text{(C.36)}
\end{aligned}$$

where we have used (C.35). Then the same reasoning as that of (C.34) which uses Jensen's inequality and (C.36) gives us the desired result Q.E.D

Proof of Lemma C.14: Let

$$\begin{aligned}
& V_{3q,aa'}(\tau) \\
\equiv\ & \sum_{m=1}^{q}\sum_{l=1}^{q} a_n(m)a_n(l)\int E\left[Z_{q,\tau,a}Z_{q,\tau,a'}\psi_{q,\tau-m}(v)\psi_{q,\tau-l}(u)\right] \\
& \times\sum_{s=2}^{\tau-2q-1}\sum_{0<s-c\le 2q} Z_{q,s,a}\psi_{q,s-m}(v)Z_{q,c,a'}\psi_{q,c-l}(u)dW(v)dW(u)
\end{aligned}$$

Then from (C.32), $V_{3q}(\tau) = a, a' = iand(i,j), i, j = 1, \cdots d \sum \sum V_{3q,aa'}(\tau)$. By the same reasoning as that for the proof of Lemma A.13, it is sufficient to show that

$E\left|\sum_{\tau=2q+2}^{n} V_{3q,aa'}(\tau)\right|^2 = O(qp^4/n)$. This follows from Minkowski's inequality and

$$
\begin{aligned}
& E\left|V_{3q,aa'}(\tau)\right|^2 \\
\leq & \left\{\sum_{m=1}^{q}\sum_{l=1}^{q} a_n(m)a_n(l)\int \left|E\left[Z_{q,\tau,a}Z_{q,\tau,a'}\psi_{q,\tau-m}(v)\psi_{q,\tau-l}(u)\right]\right|\right. \\
& \left.\times\left(\sum_{s=1}^{\tau-2q-1} E\left|Z_{q,s,a}\psi_{q,s-m}(v)\sum_{s-c\leq 2q} Z_{q,c,a'}\psi_{q,c-l}(u)\right|^2\right)^{\frac{1}{2}} dW(v)dW(u)\right\}^2 \\
\leq & \ 2C\tau q\left[\sum_{m=1}^{q} a_n(m)\right]^4 = O(\tau qp^4/n)
\end{aligned}
$$

This finishes the proof of Lemma C.14. Q.E.D.

Proof of Lemma C.15: Let

$$
\begin{aligned}
& V_{4q,aa'}(\tau) \\
\equiv & \sum_{m=1}^{q}\sum_{l=1}^{q} a_n(m)a_n(l)\int E\left[Z_{q,\tau,a}Z_{q,\tau,a'}\psi_{q,\tau-m}(v)\psi_{q,\tau-l}(u)\right] \\
& \times\sum_{s=2}^{\tau-2q-1}\sum_{s-c>2q} Z_{q,s,a}\psi_{q,s-m}(v)Z_{q,c,a'}\psi_{q,c-l}(u)dW(v)dW(u)
\end{aligned}
$$

Then from (C.32), $V_{4q}(\tau) = a, a^{'} = iand(i,j), i, j = 1, \cdots d\sum\sum V_{4q,aa'}(\tau)$. By the same reasoning as that for the proof of Lemma C.13, it is sufficient to show that $E\left|\sum_{\tau=2q+2}^{n} V_{4q,aa'}(\tau)\right|^2 = O(p)$. This follows from Minkowski's inequality, $p \to \infty$ and

$$
\begin{aligned}
& E\left|V_{4q,aa'}(\tau)\right|^2 \\
\leq & \ E\left|\sum_{m=1}^{q}\sum_{l=1}^{q} a_n(m)a_n(l)\int E\left[Z_{q,\tau,a}Z_{q,\tau,a'}\psi_{q,\tau-m}(v)\psi_{q,\tau-l}(u)\right]\right. \\
& \left.\times\sum_{s=2q+2}^{\tau-2q-1} Z_{q,s,a}\psi_{q,s-m}(v)\sum_{c=1}^{s-2q-1} Z_{q,c,a'}\psi_{q,c-l}(u)dW(v)dW(u)\right|^2
\end{aligned}
$$

$$
\begin{aligned}
\leq & \sum_{m_1=1}^{q}\sum_{m_2=1}^{q}\sum_{l_1=1}^{q}\sum_{l_2=1}^{q} a_n(m_1)a_n(m_2)a_n(l_1)a_n(l_2) \\
& \times \int\int\int\int E\left[Z_{q,0,a}Z_{q,0,a'}\psi_{q,-m_1}(v_1)\psi_{q,-l_1}(v_2)\right] \\
& \times E\left[Z_{q,0,a}Z_{q,0,a'}\psi^*_{q,-m_2}(u_1)\psi^*_{q,-l_2}(u_2)\right]\sum_{s=2q+2}^{\tau-2q-1} E\left[Z_{q,s,a}Z_{q,s,a'}\psi_{q,s-m_1}(v_1)\psi_{q,s-m_2}(u_2)\right] \\
& \times \sum_{c=1}^{s-2q-1} E\left[Z_{q,c,a}Z_{q,c,a'}\psi^*_{q,c-l_1}(v_2)\psi^*_{q,c-l_2}(u_2)\right] dW(v_1)dW(v_2)dW(u_1)dW(u_2) \\
= & \; O(\tau^2 p/n^4)
\end{aligned}
$$

by Assumption 4.4.2 and 4.4..8 that imply

$$
\begin{aligned}
& \sum_{m=1}^{\infty}|\sigma_m(v,u)| \\
\leq & \; C\sum_{m=1}^{\infty}\sum_{l=1}^{\infty} E\left|\left(Z_{0,a}Z_{0,a'}-\sigma(a,a')\right)\psi_{-m}(v)\psi_{-l}(u)\right|
\end{aligned}
$$

This finishes the proof of Lemma C.15. Q.E.D.

Proof of Theorem 4.4.2. It is sufficient to prove the following Theorems C.4 and C.5. Q.E.D.

Theorem C.4. Under the conditions of Theorem 4.4.2, $(p^{\frac{1}{2}}n)\left[\widehat{M}_0(p)-M_0(p)\right]\to^p 0$.

Theorem C.5. Under the conditions of Theorem 4.4.2 and for $a=i$ and ij, $i,j=1,\cdots d$,

$$
\begin{aligned}
& (p^{\frac{1}{2}}/n)M_0(p) \\
\to & \; ^p\left[2D\int_0^{\infty}k^4(z)dz\right]^{-1/2} \\
& \times\int\int_{-\pi}^{\pi}\left|f_a^{(0,1,0)}(\omega,0,v)-f_{0,a}^{(0,1,0)}(\omega,0,v)\right|^2 d\omega dW(v) \qquad \text{(C.37)}
\end{aligned}
$$

(C.37)

and therefore

$$(p^{\frac{1}{2}}/n)M_0(p)$$
$$\to^p \left[2D\int_0^\infty k^4(z)dz\right]^{-1/2}\int\int_{-\pi}^{\pi}\left\|f^{(0,1,0)}(\omega,0,v)-f_0^{(0,1,0)}(\omega,0,v)\right\|^2 d\omega dW(v)$$

Proof of Theorem C.4. It suffices to show that

$$n^{-1}\int\sum_{m=1}^{n}k^2(m/p)n_m\left[\left\|\widehat{\sigma}_m^{(1,0)}(0,v)\right\|^2-\left\|\widetilde{\sigma}_m^{(1,0)}(0,v)\right\|^2\right]dW(v)\to^p 0 \qquad \text{(C.38)}$$

$p^{-1}\left[\widehat{C}_0(p)-\widetilde{C}_0(p)\right] = O_p(1)$, and $p^{-1}\left[\widehat{D}_0(p)-\widetilde{D}_0(p)\right] = o_p(1)$, where $\widetilde{C}_0(p)$ and $\widetilde{D}_0(p)$ are defined in the same way as $\widehat{C}_0(p)$ and $\widehat{D}_0(p)$ in (4.28) with $\{Z_\tau\}_{\tau=1}^n$ replacing $\{\widehat{Z}_\tau\}_{\tau=1}^n$. I focus on the proof of (C.38) to save space; the proofs for $p^{-1}\left[\widehat{C}_0(p)-\widetilde{C}_0(p)\right] = O_p(1)$, and $p^{-1}\left[\widehat{D}_0(p)-\widetilde{D}_0(p)\right] = o_p(1)$ are straightforward. Because (C.5) implies that

$$n^{-1}\int\sum_{m=1}^{n}k^2(m/p)n_m\left[\left\|\widehat{\sigma}_m^{(1,0)}(0,v)\right\|^2-\left\|\widetilde{\sigma}_m^{(1,0)}(0,v)\right\|^2\right]dW(v)$$
$$= \sum\sum_{a,a'=i,(i,j);i,j=1,\cdots d} n^{-1}\int\sum_{m=1}^{n}k^2(m/p)n_m\left[\left|\widehat{\sigma}_{m,a}^{(1,0)}(0,v)\right|^2-\left|\widetilde{\sigma}_{m,a}^{(1,0)}(0,v)\right|^2\right]dW(v),$$

then it suffices to prove that

$$n^{-1}\int\sum_{m=1}^{n}k^2(m/p)n_m\left[\left|\widehat{\sigma}_{m,a}^{(1,0)}(0,v)\right|^2-\left|\widetilde{\sigma}_{m,a}^{(1,0)}(0,v)\right|^2\right]dW(v)\to^p 0 \qquad \text{(C.39)}$$

for $a = i$ and (i,j), $i,j = 1,\cdots d$. We shall show this only for the case $a = i$ with $i = 1,\cdots d$; the proofs for all other cases are similar.

Since the proof of Theorem C.5 does not depend on Theorem C.4, it follows from (C.37) that

$$n^{-1}\int\sum_{m=1}^{n-1}k^2(m/p)n_m\left|\widetilde{\sigma}_{m,i}^{(1,0)}(0,v)\right|^2 dW(v)=O_p(1) \tag{C.40}$$

for $i=1,\cdots d$.

By (C.11), (C.40), (C.8) and Cauchy-Schwartz inequality, it is sufficient for (C.39) to prove that $n^{-1}\widehat{A}_1=o_p(1)$, where $\widehat{A}_1$ is defined as in (C.8). Then (C.9) implies further that it is enough to show

$$n^{-1}\int\sum_{m=1}^{n-1}k^2(m/p)n_m\left|\widehat{B}_{hm}(v)\right|^2 dW(v)=o_p(1)$$

for $h=1,\cdots,6$.

I first consider the case $h=1$. By Cauchy-Schwartz inequality and $\left|\widehat{\delta}_\tau(v)\right|\leqq 2$,

$$\begin{aligned}\left|\widehat{B}_{1m}(v)\right|^2 &\leqq \left[n_m^{-1}\sum_{\tau=m+1}^{n}\left(\widehat{Z}_{\tau,i}-Z_{\tau,i}\right)^2\right]\left[n_m^{-1}\sum_{\tau=m+1}^{n}\left|\widehat{\delta}_\tau(v)\right|^2\right]\\ &\leqq n_m^{-1}\sum_{\tau=1}^{n}\left(\widehat{Z}_{\tau,i}-Z_{\tau,i}\right)^2\end{aligned}$$

Then it follows from (C.3) and (C.11) and Assumption A.6 that

$$\begin{aligned}&n^{-1}\int\sum_{m=1}^{n-1}k^2(m/p)n_m\left|\widehat{B}_{1m}(v)\right|^2 dW(v)\\ \leqq&\left[\sum_{\tau=1}^{n}\left(\widehat{Z}_{\tau,i}-Z_{\tau,i}\right)^2\right]\sum_{\tau=m+1}^{n}a_n(m)\left[\int dW(v)\right]^2=O_p(p/n)\end{aligned}$$

The proof for case $h=2$ is similar, noting that

$$\left|n_m^{-1}\sum_{\tau=m+1}^{n}\left(\widehat{Z}_{\tau,i}-Z_{\tau,i}\right)\right|^2\leqq n_m^{-1}\sum_{\tau=m+1}^{n}\left|\widehat{Z}_{\tau,i}-Z_{\tau,i}\right|^2$$

Next consider the case $h=3$. Still by the Cauchy-Schwartz inequality, I have

$$\begin{aligned}\left|\widehat{B}_{3m}(v)\right|^2 &\leqq \left(n_m^{-1}\sum_{\tau=1}^{n}Z_{\tau,i}^2\right)n_m^{-1}\sum_{\tau=m+1}^{n}\left|\widehat{\delta}_{\tau-m}(v)\right|^2\\ &\leqq \|v\|^2\left(n_m^{-1}\sum_{\tau=1}^{n}Z_{\tau,i}^2\right)n_m^{-1}\sum_{\tau=m+1}^{n}\left(\widehat{Z}_{\tau,i}-Z_{\tau,i}\right)^2\end{aligned}$$

It then follows that

$$
\begin{aligned}
& n^{-1} \int \sum_{m=1}^{n-1} k^2(m/p) n_m \left| \widehat{B}_{3m}(v) \right|^2 dW(v) \\
\leqq \; & \left(n^{-1} \sum_{\tau=1}^{n} Z_{\tau,i}^2 \right) \left[n^{-1} \sum_{\tau=1}^{n} \left(\widehat{Z}_{\tau,i} - Z_{\tau,i} \right)^2 \right] \sum_{\tau=1}^{n-1} k^2(m/p) \int \|v\|^2 dW(v) \\
= \; & O_p(p/n)
\end{aligned}
$$

The proof for the cases $h = 4, 5, 6$ is similar to the case $h = 3$, noting that

$$
\left| n_m^{-1} \sum_{\tau=m+1}^{n} \widehat{\delta}_\tau(v) \right|^2 \leqq n_m^{-1} \sum_{\tau=m+1}^{n} \left| \widehat{\delta}_\tau(v) \right|^2
$$

This completes the proof of Theorem C.4. Q.E.D.

Proof of Theorem C.5. The proof is a straightforward extension for that of Hong(1999, Proof of Theorem 5), for the case $(m, l) = (1, 0)$ and $W_1(\cdot) = \delta(\cdot)$, the Dirac delta function. I omit it here to save space. Note that Assumption 4.4.8 is needed here. Q.E.D.

Proof of Theorem 4.4.3. It is sufficient to prove the Theorems C.6 and C.7 below. Q.E.D.

Theorem C.6. Under the conditions of Theorem 4.4.3, $\widehat{M}_0(\widehat{p}) - M_0(\widehat{p}) = o_p(1)$.

Theorem C.7. Under the conditions of Theorem 4.4.3, $M_0(\widehat{p}) - M_0(p) = o_p(1)$.

Proof of Theorem C.6. Define

$$
\widehat{B} \equiv \sum_{m=1}^{n-1} k^2(m/\widehat{p}) n_m \int \left[\left\| \widehat{\sigma}_m^{(1,0)}(0, v) \right\|^2 - \left\| \widetilde{\sigma}_m^{(1,0)}(0, v) \right\|^2 \right] dW(v)
$$

Then it suffices to show that $p^{-\frac{1}{2}} \widehat{B} = o_p(1)$, $p^{-\frac{1}{2}} \left[\widehat{C}_0(\widehat{p}) - \widetilde{C}_0(\widehat{p}) \right] = o_p(1)$, and $p^{-1} \left[\widehat{D}_0(\widehat{p}) - \widetilde{D}_0(\widehat{p}) \right] = o_p(1)$. I only show $p^{-\frac{1}{2}} \widehat{B} = o_p(1)$ here to save space; the proof of the other two is similar. Since

$$\begin{aligned}
\widehat{B} &= \sum_{a=i,(i,j);i,j=1,\cdots d} \left\{ \sum_{m=1}^{n-1} k^2(m/\widehat{p}) n_m \int \left[\left|\widehat{\sigma}_{m,a}^{(1,0)}(0,v)\right|^2 - \left|\widetilde{\sigma}_{m,a}^{(1,0)}(0,v)\right|^2 \right] dW(v) \right\} \\
&\equiv \sum_{a=i,(i,j);i,j=1,\cdots d} \widehat{B}_a
\end{aligned}$$

then it is sufficient to show $p^{-\frac{1}{2}}\widehat{B}_a = o_p(1)$ for $a = i$ and ij, $i, j = 1, \cdots d$. We shall only show this holds for $a = i$; the proofs for all other cases are similar.

By the conditions on $k(\cdot)$ implied by Assumption 4.4.5, there exists a symmetric monotonic decreasing function $k_0(z)$ for $z > 0$ such that $|k(z)| \leqq k_0(z)$ for all $z > 0$ and $k_0(\cdot)$ satisfies Assumption 4.4.5. Then for any constants $\epsilon, \eta > 0$,

$$\begin{aligned}
& P\left(p^{-\frac{1}{2}}\left|\widehat{B}_i\right| > \epsilon\right) \\
\leq\ & P\left(p^{-\frac{1}{2}}\left|\widehat{B}_i\right| > \epsilon, \left|\widehat{p}/p - 1\right| \leq \eta\right) + P\left(\left|\widehat{p}/p - 1\right| > \eta\right)
\end{aligned}$$

where the second term vanishes asymptotically for all $\eta > 0$ and given $\widehat{p}/p - 1 \to^p 0$. Therefore, by the definition of convergence in probability, it remains to show that the first terms also vanishes as $n \to \infty$.

Because $\left|\widehat{p}/p - 1\right| \leq \eta$ implies $\widehat{p} \leq (1+\eta)p$, for $\left|\widehat{p}/p - 1\right| \leq \eta$

$$\begin{aligned}
& p^{-\frac{1}{2}}\left|\widehat{B}_i\right| \\
\leq\ & (1+\eta)^{\frac{1}{2}} \left[(1+\eta)p\right]^{-\frac{1}{2}} \sum_{m=1}^{n-1} k_0^2 \left[m/(1+\eta)p\right] \\
& \times n_m \left[\left|\widehat{\sigma}_{m,i}^{(1,0)}(0,v)\right|^2 - \left|\widetilde{\sigma}_{m,i}^{(1,0)}(0,v)\right|^2 \right] \\
\to\ & {}^p 0
\end{aligned}$$

for any $\eta > 0$ given (C.9), where the inequality follows from $|k(z)| \leqq k_0(z)$ for all $z > 0$. This completes the proof of Theorem C.6. Q.E.D.

Proof of Theorem C.7. The proof is a straightforward extension from those of Theorem A.7 of Hong and Lee(2005) which follows a reasoning analogous to the proof of Hong(1999, Theorem4). Note that the latter uses an assumption which is exactly the same as Assumption 4.4.10. That is, Assumption 4.4.10 is also used in this proof. Q.E.D.

APPENDIX D

APPENDIX OF CHAPTER 5

Throughout the Appendix, $C \in (1, \infty)$ denotes a generic bounded constant.

Proof of Theorem 5.4.1. By the discussions in Section 5.2, $\int |H(\theta, x)|^2 \, dP_{X_{\tau-1}}(x)$ has a unique minimum at θ_0. Then by the standard theory of M-estimators, we only have to show that

$$\int |H_{n-1}(\theta, x)|^2 \, dP_{n-1}(x) \to^{a.s.} \int |H(\theta, x)|^2 \, dP_{X_{\tau-1}}(x) \tag{D.1}$$

uniformly in θ, where $P_{n-1}(x) = \frac{1}{n-1} \sum_{\tau=2}^{n} 1(X_{\tau-1} = x)$ is the empirical analog of $P_{X_{\tau-1}}(x)$. By Continuous Mapping theorem, it is sufficient for (D.1) to show

$$H_{n-1}(\theta, x) \to^{a.s.} H(\theta, x) \tag{D.2}$$

uniformly in (x, θ). Let $H^a(\theta, x)$ and $H^a_{n-1}(\theta, x)$ be the a-th components of $H(\theta, x)$ and $H_{n-1}(\theta, x)$ respectively, i.e.,

$$H^a(\theta, x) = E\left[Z^a_\tau(\theta) 1 (X_{\tau-1} \le x)\right]$$

and

$$H^a_{n-1}(\theta, x) = (n-1)^{-1} \sum_{\tau=2}^{n} Z^a_\tau(\theta) 1 (X_{\tau-1} \le x)$$

for $a = i, ij$ with $i, j = 1, \cdots, d$. Then to prove (D.2), we only have to show

$$H^a_{n-1}(\theta, x) \to^{a.s.} H^a(\theta, x) \tag{D.3}$$

uniformly in (x, θ), which is similar to the proof in Domínguez and Lobato (2004, Theorem 1) and follows from Ranga Rao (1962). Q.E.D.

Proof of Theorem 5.4.2. We shall prove the result for the one-dimensional semi-parametric diffusion model. Those for the general multivariate cases are just straightforward extensions. Similar to the proof of Theorem 5.4.1 and by Theorem 2.1 of Newey and McFadden (1994), it is sufficient to show that

$$\begin{aligned} H^1_{n-1}\left(\theta, x, \widehat{[X, X]}\right) &\rightarrow {}^p H^1\left(\theta, x, [X, X]\right) \\ H^2_{n-1}\left(\theta, x, \widehat{[X, X]}\right) &\rightarrow {}^p H^2\left(\theta, x, [X, X]\right) \end{aligned} \tag{D.4}$$

both uniformly in (x, θ), where

$$\begin{aligned} H^1\left(\theta, x, [X, X]\right) &= E\left[Z^x_\tau(\theta) 1\left(X_{\tau-1} \leq x\right)\right] \\ H^2\left(\theta, x, [X, X]\right) &= E\left[Z^{x^2}_\tau(\theta, [X, X]) 1\left(X_{\tau-1} \leq x\right)\right] \end{aligned}$$

and correspondingly

$$\begin{aligned} H^1_{n-1}\left(\theta, x, \widehat{[X, X]}\right) &= (n-1)^{-1} \sum_{\tau=2}^{n} Z^x_\tau(\theta) 1\left(X_{\tau-1} \leq x\right) \\ H^2_{n-1}\left(\theta, x, \widehat{[X, X]}\right) &= (n-1)^{-1} \sum_{\tau=2}^{n} Z^{x^2}_\tau(\theta, \widehat{[X, X]}) 1\left(X_{\tau-1} \leq x\right) \end{aligned}$$

The proof of the first relation in (D.4) is exactly the same as that of (D.3) and it is actually an *a.s.*convergence which implies the convergence in probability. Next, let $H^2_{n-1}\left(\theta, x, [X, X]\right)$ be defined as $H^2_{n-1}\left(\theta, x, \widehat{[X, X]}\right)$ above with $[X, X]$ replacing $\widehat{[X, X]}$. Still similar to (D.3), we have

$$H^2_{n-1}\left(\theta, x, [X, X]\right) \rightarrow^p H^2\left(\theta, x, [X, X]\right) \tag{D.5}$$

uniformly in (x,θ), which can also be *a.s.*convergence in fact. By (D.5), it is sufficient for the second relation in (D.4) to prove

$$H^2_{n-1}\left(\theta, x, \widehat{[X,X]}\right) - H^2_{n-1}\left(\theta, x, [X,X]\right) = o_P(1) \tag{D.6}$$

It follows from the definitions of $H^2_{n-1}\left(\theta, x, \widehat{[X,X]}\right)$ and $H^2_{n-1}\left(\theta, x, [X,X]\right)$ that

$$\begin{aligned}
& H^2_{n-1}\left(\theta, x, \widehat{[X,X]}\right) - H^2_{n-1}\left(\theta, x, [X,X]\right) \\
= \;& (n-1)^{-1}\sum_{\tau=2}^{n}\left[Z^{x^2}_{\tau}(\theta, \widehat{[X,X]}) - Z^{x^2}_{\tau}(\theta, [X,X])\right] 1\left(X_{\tau-1} \le x\right) \\
\le \;& (n-1)^{-1}\sum_{\tau=2}^{n}\left(\widehat{[X,X]}^{\tau\Delta}_{(\tau-1)\Delta} - [X,X]^{\tau\Delta}_{(\tau-1)\Delta}\right) \\
= \;& O_p(h^{1/2}) = o_p(1)
\end{aligned}$$

where (5.25) is used in the second equality. This completes the proof. Q.E.D.

Proof of Theorem 5.4.3: This is just a simple application of Theorem 5.4.1 since $\widehat{\theta}_{SI}$ is a special case of $\widehat{\theta}_0$. Q.E.D.

Proof of Theorem 5.4.4: The first order conditions of the minimization problem (5.22) are

$$(n-1)^{-1}\sum_{l=2}^{n} 2\left[\cdot H_{n-1}(\widehat{\theta}_0, X_l)' H_{n-1}(\widehat{\theta}_0, X_l)\right] = 0 \tag{D.7}$$

Assumption 5.5.5 and the mean value theorem imply that

$$H_{n-1}(\widehat{\theta}_0, X_l) = H_{n-1}(\theta_0, X_l) + \cdot H_{n-1}(\theta^{\dagger}, X_l)\left(\widehat{\theta}_0 - \theta_0\right)$$

where $\theta^{\dagger}$ is between $\widehat{\theta_0}$ and θ_0 and its value may be different for each row in the matrix $\cdot H_{n-1}(\theta^{\dagger}, X_l)$. Plugging this back into (D.7) yields

$$(n-1)^{-1} \sum_{l=2}^{n} \cdot H_{n-1}(\widehat{\theta_0}, X_l)' \left[H_{n-1}(\theta_0, X_l) + \cdot H_{n-1}(\theta^{\dagger}, X_l) \left(\widehat{\theta_0} - \theta_0 \right) \right] = 0$$

which further implies that

$$\begin{aligned} & \sqrt{n} \left(\widehat{\theta_0} - \theta_0 \right) \\ = \; & - \left[(n-1)^{-1} \sum_{l=2}^{n} \cdot H_{n-1}(\widehat{\theta_0}, X_l)' \cdot H_{n-1}(\theta^{\dagger}, X_l) \right]^{-1} \\ & \times \left[(n-1)^{-1} \sum_{l=2}^{n} \cdot H_{n-1}(\widehat{\theta_0}, X_l)' \sqrt{n} H_{n-1}(\theta_0, X_l) \right] \end{aligned}$$

Then, the result follows from the Continuous Mapping theorem and Propositions D.1 and D.2 below. Q.E.D.

Proposition D.1: Under the conditions of Theorem 5.4.4,

$$(n-1)^{-1} \sum_{l=2}^{n} \cdot H_{n-1}(\widehat{\theta_0}, X_l)' \cdot H_{n-1}(\theta^{\dagger}, X_l) \to^{p} \mathcal{M}_q.$$

Proposition D.2: Under the conditions of Theorem 5.4.4,

$$(n-1)^{-1} \sum_{l=2}^{n} \cdot H_{n-1}(\widehat{\theta_0}, X_l)' \sqrt{n} H_{n-1}(\theta_0, X_l) \to^{d} N(0, \Omega_q).$$

Proof of Proposition D.1: Similar to (D.2), we have $\cdot H_{n-1}(\theta, x) \to^{a.s.} \cdot H_{n-1}(\theta, x)$ uniformly in (θ, x). Then the proof is done by Continuous Mapping theorem and the fact that $\widehat{\theta_0} \to^{a.s.} \theta_0$ and $\theta^{\dagger} \to^{a.s.} \theta_0$ which follows from Theorem 5.4.1. Q.E.D.

Proof of Proposition D.2: This result follows from Continuous Mapping theorem, $\cdot H_{n-1}(\theta, x) \to^{a.s.} \cdot H_{n-1}(\theta, x)$ uniformly in (θ, x) which is similar to (D.2) and Lemma D.1 below combined with the fact that the integrated weighted

Gaussian process follows a normal distribution(see, for instance, Tanaka (1996, Ch.2)). Q.E.D.

Lemma D.1: Under the conditions of Theorem 5.4.3,

$$\sqrt{n}H_{n-1}(\theta_0,\cdot) \Longrightarrow B_{\Gamma}(\cdot)$$

where B_{Γ} denotes a d'-dimensional centered Gaussian vector process in the product space $\left(D\left[\mathbb{R}\right]^d\right)^{d'} = d'\underbrace{D\left[\mathbb{R}\right]^d \times \cdots D\left[\mathbb{R}\right]^d}$, $\Longrightarrow$ denotes weak convergence in $\left(D\left[\mathbb{R}\right]^d\right)^{d'}$, $D\left[\mathbb{R}\right]^d$ is the natural extension of $D\left[0,1\right]^d$ in the sense of Stute (1997), $D\left[0,1\right]^d$ is defined in Bickel and Wichura (1971) as an extension of $D\left[0,1\right]$ in Billingsley(1999, Ch.3). The covariance structure of the process B_{Γ} is given by Γ as defined in Theorem 5.4.3.

Proof of Lemma D.1: Although $D\left[\mathbb{R}\right]^d$, as a natural extension of $D\left[0,1\right]^d$, is discussed and used in Domínguez and Lobato (2004), its multivariate counterpart has not been considered. Therefore, we shall define the product space $\left(D\left[\mathbb{R}\right]^d\right)^{d'}$ and discuss both its topological and measurable properties and related convergence concepts. Note that here we only extend $D\left[\mathbb{R}\right]^d$ to its finite dimensional product; for the extension to the infinite dimensional product, see Ledoux and Talagrand (1991) for reference. Following the same procedure as that of Davidson (1994, 27.7) who extends $C\left[0,1\right]$ to its m-dimensional product $\left(C\left[0,1\right]\right)^m$, we can endow $\left(D\left[\mathbb{R}\right]^d\right)^{d'}$ with the metric $\lambda_{d'}(x,y) \equiv \max_{e=1,\cdots,d'}\left\{\lambda\left(x_e,y_e\right)\right\}$ where $x,y \in \left(D\left[\mathbb{R}\right]^d\right)^{d'}$ and $\lambda\left(\cdot,\cdot\right)$ is the metric on $D\left[\mathbb{R}\right]^d$ making $D\left[\mathbb{R}\right]^d$ complete and separable. Then the metric $\lambda_{d'}\left(\cdot,\cdot\right)$ will induce the product topology on $\left(D\left[\mathbb{R}\right]^d\right)^{d'}$ and the coordinate projections remain continuous. Since $D\left[\mathbb{R}\right]^d$ is separable, then $\left(D\left[\mathbb{R}\right]^d\right)^{d'}$ is also separable by Davidson(1994, 6.16). Let $\mathcal{B}_D$ and $\mathcal{B}_D^{d'}$ be the Borel σ-fields for $D\left[\mathbb{R}\right]^d$ and $\left(D\left[\mathbb{R}\right]^d\right)^{d'}$ respectively, i.e., the σ-fields generated by the open sets of $D[\mathbb{R}]^d$ and $\left(D[\mathbb{R}]^d\right)^{d'}$ respectively. Then by Theorem 1.10 of Parthasarathy (1967) as well as the separability of $D[\mathbb{R}]^d$, $\mathcal{B}_D^{d'} = d'\underbrace{\mathcal{B}_D \otimes \cdots \otimes \mathcal{B}_D}$ where the right hand side is the product σ-field defined as $\sigma\{A_1 \times \cdots \times A_{d'} : A_e \in \mathcal{B}_D$ for any $e = 1, \cdots, d'\}$. Therefore, $\left(\left(D[\mathbb{R}]^d\right)^{d'}, \mathcal{B}_D^{d'}\right)$ is a measurable space for which all the standard weak convergence related concepts and results for metric space apply(See Billingsley (1995, Ch.1)).

By similar arguments as those for proving weak convergence in $D[\mathbb{R}]^d$(See Domínguez and Lobato (2004) and Bickel and Wichura (1971)), we need to show the finite-dimensional distributions of the process $\sqrt{n}H_{n-1}(\theta_0, \cdot)$ are asymptotically normal with the approximate covariance structure and that the process $\sqrt{n}H_{n-1}(\theta_0, \cdot)$ is tight. Denote the components of $Z_\tau(\theta)$ as $Z_\tau^a(\theta)$ for $a = 1, \cdots, d'$ and corresponding $\sqrt{n}H_{n-1}^a(\theta_0, \cdot) = \sqrt{n}(n-1)^{-1}\sum_{\tau=2}^n Z_\tau^a(\theta)1(X_{\tau-1} \leq x)$. Then convergence of finite-dimensional distributions refers to the weak convergence of vectors of the form

$$\left(\sqrt{n}H_{n-1}^1(\theta_0, x_1^1), \cdots, \sqrt{n}H_{n-1}^1(\theta_0, x_{p_1}^1), \cdots, \sqrt{n}H_{n-1}^{d'}(\theta_0, x_1^{d'}), \cdots, \sqrt{n}H_{n-1}^{d'}(\theta_0, x_{p_{d'}}^{d'})\right)$$

for arbitrary $p_1, \cdots, p_{d'} \in \mathbb{N}$ and $x_{p_a}^a \in \mathbb{R}^d$ for $a = 1, \cdots, d'$. This result can be obtained by a straightforward extension of Corollary 3.1 in Hall and Heyde (1980) and the Cramer-Wold theorem. Note that the *m.d.s.*property of $Z_\tau(\theta)$ is used here. Next, we consider the tightness of $\sqrt{n}H_{n-1}(\theta_0, \cdot)$. By Davidson (1994, 26.23), it is sufficient to show that $\sqrt{n}H_{n-1}^a(\theta_0, \cdot)$ is tight in $D[\mathbb{R}]^d$. This can be proved by exactly the same reasoning as the proof of Lemma 2 in Domínguez and Lobato(2004) and the proof is finished. Q.E.D.

Proof of Theorem 5.4.5: Similar to the proof of Theorem 5.4.2, we shall only prove the result for the one-dimensional semi-parametric diffusion model. Those for the general multivariate cases are just straightforward extensions. Let $\widetilde{\theta}_{RV} = \arg\min_{\theta} \frac{1}{n-1} \sum_{l=2}^{n} \left|\widetilde{H}_{n-1}(\theta, X_l)\right|^2$ where $\widetilde{H}_{n-1}(\theta, x)$ is defined in the same way as $H_{n-1}\left(\theta, x, \widehat{[X,X]}\right)$ with $[X,X]$ replacing $\widehat{[X,X]}$. Then $\widetilde{\theta}_{RV}$ is a special case of $\widehat{\theta}_0$ and hence $\sqrt{n}\left(\widetilde{\theta}_{RV} - \theta_0\right) \to^d N(0, \mathcal{V}_{RV})$ by Theorem 5.4.4. The proof will be finished if we can show $\sqrt{n}\left(\widehat{\theta}_{RV} - \widetilde{\theta}_{RV}\right) = o_p(1)$. By the Continuous Mapping theorem, it is sufficient to show

$$\left|H_{n-1}\left(\theta, X_l, \widehat{[X,X]}\right)\right|^2 - \left|\widetilde{H}_{n-1}(\theta, X_l)\right|^2 = o_p(n^{-1/2})$$

uniformly in θ, for each $l = 2, \cdots, n$. By (5.42) and the definition of $\widetilde{H}_{n-1}(\theta, x)$,

$$\begin{aligned}
& \left|\left|H_{n-1}\left(\theta, X_l, \widehat{[X,X]}\right)\right|^2 - \left|\widetilde{H}_{n-1}(\theta, X_l)\right|^2\right| \\
= & \left|\left(\frac{1}{n-1}\sum_{\tau=2}^{n} Z_\tau^{x}(\theta, \widehat{[X,X]})1\left(X_{\tau-1} \le X_l\right)\right)^2 + \left(\frac{1}{n-1}\sum_{\tau=2}^{n} Z_\tau^{x^2}(\theta, \widehat{[X,X]})1\left(X_{\tau-1} \le X_l\right)\right)^2\right. \\
& \left. -\left(\frac{1}{n-1}\sum_{\tau=2}^{n} Z_\tau^{x}(\theta)1\left(X_{\tau-1} \le X_l\right)\right)^2 - \left(\frac{1}{n-1}\sum_{\tau=2}^{n} Z_\tau^{x^2}(\theta)1\left(X_{\tau-1} \le X_l\right)\right)^2\right| \\
= & \left|\left(\frac{1}{n-1}\sum_{\tau=2}^{n} Z_\tau^{x^2}(\theta, \widehat{[X,X]})1\left(X_{\tau-1} \le X_l\right)\right)^2 - \left(\frac{1}{n-1}\sum_{\tau=2}^{n} Z_\tau^{x^2}(\theta)1\left(X_{\tau-1} \le X_l\right)\right)^2\right| \\
= & \left|\frac{2}{n-1}\sum_{\tau=2}^{n} Z_\tau^{x^2}(\theta)\right| \left[\frac{1}{n-1}\sum_{\tau=2}^{n}\left(Z_\tau^{x^2}(\theta, \widehat{[X,X]}) - Z_\tau^{x^2}(\theta)\right)\right]^2 \\
\equiv & \; U_1 \times U_2
\end{aligned} \tag{D.8}$$

We consider the first term in (D.8). By Assumption 5.5.1 and 5.5.2, an application of Ergodic theorem (see Durrett (2005, Ch.6)) implies that $\frac{2}{n-1}\sum_{\tau=2}^{n} Z_\tau^{x^2}(\theta) \to^{a.s.} E\left|Z_\tau^{x^2}(\theta)\right| < E[Y]$. Therefore,

$$U_1 \leq C a.s. \tag{D.9}$$

uniformly in θ. Then it follows from (5.40) that

$$\begin{aligned} Z_\tau^{x^2}(\theta, [\widehat{X, X}]) - Z_\tau^{x^2}(\theta) &= [X, X]_{(\tau-1)\Delta}^{\tau\Delta} - [\widehat{X, X}]_{(\tau-1)\Delta}^{\tau\Delta} \\ &= O_p(h^{1/2}), \end{aligned}$$

uniformly in θ, where the second equality is based on (5.25) which follows from Barndorff-Nielsen and Shephard (2004) and Bandi and Russell (2005) and the uniformity in θ is obvious since $Z_\tau^{x^2}(\theta, [\widehat{X, X}]) - Z_\tau^{x^2}(\theta)$ does not involve θ at all.

$$U_2 = \left[O_p(h^{1/2})\right]^2 = O_p\left(n^{-1/2-\alpha}\right) = o_p(n^{-1/2}) \tag{D.10}$$

uniformly in θ. The desired result is then proved by combing (D.8)-(D.10). Q.E.D.

Proof of Theorem 5.4.6: This is just a simple application of Theorem 5.4.4 since $\widehat{\theta}_{SI}$ is a special case of $\widehat{\theta}_0$. Q.E.D.

BIBLIOGRAPHY

Ahn, D., and B. Gao, 1999, “A Parametric Nonlinear Model of Term Structure Dynamics,” Review of Financial Studies, 12, 721–762.

Ahn, C.M., Thompson, H.E., (1988). ”Jump-diffusion processes and the term structure of interest rates”. Journal of Finance 43, 155–174.

Ahn, D.-H., R. F. Dittmar, and A. R. Gallant, (2002), “Quadratic Term Structure Models: Theory and Evidence,” Review of Financial Studies, 15, 243-288.

Ahn, D.-H, R.F. Dittmar, A. R. Gallant, and B. Gao, (2003), ”Purebred or Hybrid?: Reproducing the Volatility in Terms Structure Dynamics”, Journal of Econometrics, 116, 147-180.

Ait-Sahalia, Y. (1996a), ”Testing Continuous-Time Models of the Spot Interest Rate,” Review of Financial Studies, 9, 385-426.

Aït-Sahalia, Y. (1999). Transition densities for interest rate and other nonlinear diffusions. Journal of Finance 54, 1361—1395.

Ait-Sahalia, Y. (1996b), ”Nonparametric Pricing of Interest Rate Derivative Securities,” Econometrica, 64, 527-560.

Ait-Sahalia, Y. (2002a) ”Telling from Discrete Data Whether the Underlying Continuous-Time Model Is a Diffusion,” Journal of Finance, 57, 2075-2112.

Ait-Sahalia, Y. (2002b) ”Maximum-Likelihood Estimation of Discretely Sampled Diffusions: A Closed-Form Approach,” Econometrica, 70, 223-262.

Ait-Sahalia, Y. (2007), ”Estimating Continuous-Time Models Using Discretely Sampled Data”, in Advances in Economics and Econometrics, Theory and Applications, Ninth World Congress, Econometric Society Monographs, Cambridge University Press.

Ait-Sahalia, Y. (2008). "Closed-Form Likelihood Expansions for Multivariate Diffusions", Annals of Statistics, 36, 906-937.

Ait-Sahalia, Y., J. Fan and H. Peng (2009): "Nonparametric Transition-Based Tests for Jump-Diffusions", Journal of the American Statistical Association, 2009, 104, 1102-1116.

Ait-Sahalia Y., Hansen, L. and Scheinkman, J. 2004 "Operator Methods for Continuous-Time Markov Processes", Handbook of Financial Econometrics.

Ait-Sahalia, Y. and Jacod, J. (2008) "Testing for Jumps in a Discretely Observed Process," forthcoming in Annals of Statistics.

Aït-Sahalia, Y. and R. Kimmel (2007), "Maximum Likelihood Estimation of Stochastic Volatility Models", Journal of Financial Economics, 2007, 83, 413-452.

Aït-Sahalia, Y. and R. Kimmel, (2010). "Estimating affine multifactor term structure models using closed-form likelihood expansions", Journal of Financial Economics, Volume 98, Issue 1, 113-144

Aıt-Sahalia, Y., Mykland, P.A., (2003). "The effects of random and discrete sampling when estimating continuous-time diffusions". Econometrica 71, 483–549.

Aıt-Sahalia, Y., Mykland, P.A., (2004). "Estimating Diffusions with Discretely and Possibly Randomly Spaced Data: A General Theory", Annals of Statistics, 2004, 32, 2186-2222.

Aït-Sahalia, Y. and P. Mykland (2008). "An Analysis of Hansen-Scheinkman Moment Estimators for Discretely and Randomly Sampled Diffusions", Journal of Econometrics, 144, 1-26.

Ait-Sahalia, Y., Mykland, P. and Zhang, L. 2005, ”A Tale of Two Time Scales: Determining Integrated Volatility with Noisy High-Frequency Data”, Journal of the American Statistical Association, 100, 1394-1411.

Akgiray, V., and Booth, G., (1988), ”Mixed Diffusion-Jump Process Modeling Of Exchange Rate Movements,” Review of Economics and Statistics, November, pp. 631-637.

Anatolyev, S. (2005): “GMM, GEL, Serial correlation, and asymptotic bias,” Econometrica, 73, 983–1002.

Andersen, T., and Benzoni, L., (2010), ”Do Bonds Span Volatility Risk in the U.S. Treasury Market? A Specification Test for Affine Term Structure Models”. The Journal of Finance, 65: 603–653

Andersen, T.G., Benzoni, L. and Lund, J. 2002, ”An Empirical Investigation of Continuous-Time Models for Equity Returns,” Journal of Finance 57 (2002): 1239-1284.

Andersen, T., Benzoni, L., Lund, J., (2004). ”Stochastic volatility, mean drift, and jumps in the short-term interest rate”. Working paper, Northwestern University.

Andersen, T.G., Bollerslev, T., Diebold, F.X., Labys, P., 2003. Modeling and forecasting realized volatility. Econometrica 71, 579–625.

Andersen, T.G., Bollerslev, T. and Dobrev, D. (2007) ”No-arbitrage Semi-Martingale Restrictions for Continuous-Time Volatility Models Subject to Leverage Effects, Jumps and I.I.D. Noise: Theory and Testable Distributional Implications,” Journal of Econometrics, 138, 125-180.

Andersen, T. G., and J. Lund, (1997), “Estimating Continuous-Time Stochastic

Volatility Models of the Short-Term Interest Rate," Journal of Econometrics, 72, 343-377.

Backus, D., S. Foresi, A. Mozumdar, and L. Wu (2001). "Predictable changes in yields and forward rates," Journal of Financial Economics 59, 281-311.

Bahadur, R.R. (1960): "Stochastic Comparison of Tests," Annals of Mathematical Statistics, 31, 276-195.

Bakshi, G.S., Ju, N. and Ou-Yang H, (2006), "Estimation of Continuous-time Models with an Application to Equity Volatility," Journal of Financial Economics 82.

Ball, C. A., and W. N. Torous, 1996, Unit Roots and the Estimation of Interest Rate Dynamics, Journal of Empirical Finance, 3, 215–238.

Bandi, F. M., and R. Reno, (2008). "Nonparametric Stochastic Volatility". SSRN: http://ssrn.com/abstract=1158438.

Bandi, F. M. and Phillips, P.C.B. (2003) "Fully Nonparametric Estimation of Scholar Diffusion Models," Econometrica, 71, 241-283.

Bandi, F. M. and Phillips, P.C.B. (2007), "A Simple Approach to the Parametric Estimation of Potentially Non-stationary Diffusions.", Journal of Econometrics, 137(2), pp. 354-95.

Bandi, F. M. and Russell, J.R. 2005. "Realized covariation, realized beta, and microstructure noise", Graduate School of Business, The University of Chicago.

Bandi, F. M. and Russell, J.R. 2008. "Microstructure Noise, Realized Variance, and Optimal Sampling", Review of Economic Studies, Volume 75, Issue 2, Pages 339-369.

Bansal, R., and H. Zhou, (2002), "Term Structure of Interest Rates with Regime Shifts," Journal of Finance, 57, 1997-2043.

Bansal, R., G. Tauchen, and H. Zhou, (2004), "Regime Shifts, Risk Premiums in the Term Structure, and the Business Cycle," Journal of Business and Economic Statistics, 22, 396 – 409

Barndorff-Nielsen, O., (1998), "Processes of normal inverse Gaussian type", Finance and Stochastics, 41-68.

Barndorff-Nielsen, O.E., Shephard, N., 2002. "Econometric analysis of realized volatility and its use in estimating stochastic volatility models", Journal of The Royal Statistical Society Series B statistical Methodology, vol. 64, 253-280.

Barndorff-Nielsen, O.E., Shephard, N., 2004. Power and bipower variation with stochastic volatility and jumps. Journal of Financial Econometrics 2, 1–37.

Barndorff-Nielsen, O. E. and Shephard, N. (2006) "Econometrics of Testing for Jumps in Financial Economics Using Bipower Variation," Journal of Financial Econometrics, 4,1-30.

Bates, D., (1996), "Jumps and Stochastic Volatility: Exchange Rate Processes Implicit in Deutsche Mark Options", Review of Financial Studies 9, 69-107.

Bekaert, G., R. Hodrick, and D. Marshall, (1997), "On Biases in Tests of the Expectations Hypothesis of the Term Structure of Interest Rates," Journal of Financial Economics, 44, 309-348.

Bekaert, G., R. Hodrick, and D. Marshall, (2001), "Peso Problem Explanations for Term Structure Anomalies," Journal of Monetary Economics, 48, 241-270.

Bertoin, J., (1996), "Levy Processes". Cambridge University Press, Cambridge.

Bickel, P. J., and M. J. Wichura (1971): "Convergence Criteria for Multiparameter Stochastic Processes and Some Applications," Annals of Mathematical Statistics, 42, 1656–1670.

Bierens, H. J. (1994). Topics in Advanced Econometrics, New York: Cambridge University Press.

Bikbov R. and M. Chernov (2010). "Yield curve and volatility: Lessons from Eurodollar futures and options". Journal of Financial Econometrics, forthcoming.

Billingsley, P. (1995): "Probability and Measure". New York: Wiley and Sons.

Billingsley, P. (1999): "Convergence of Probability Measure". New York: Wiley and Sons.

Brandt, M., and P. Santa-Clara, (2002), "Simulated Likelihood Estimation of Diffusions with an Application to Exchange Rate Dynamics in Incomplete Markets," Journal of Financial Economics 63, 161–210.

Bremaud, P. (1981). "Point Processes and Queues, Martingale Dynamics". New York: Springer-Verlag.

Brillinger, D.R. and Rosenblatt, M. 1967a. Asymptotic theory of estimates of kth order spectra. In Spectral Analysis of Time Series, ed. B. Harris. New York: Wiley.

Brillinger, D.R. and Rosenblatt, M. 1967b. Computation and interpretation of the kth order spectra. In Spectral Analysis of Time Series, ed. B. Harris. New York: Wiley.

Brown, B.M. 1971, “Martingale Limit Theorems,” Annals of Mathematical Statistics, 42, 59–66.

Buraschi, A., A. Cieslak and F. Trojan, (2008), ”Correlation Risk and the Term Structure of Interest Rates”, working paper.

Campbell, J. Y., (1986) “A defense of the traditional hypotheses about the term structure of interest rates.” Journal of Finance 41, 183–193.

Campbell, J. Y. and R. J. Shiller (1991). ”Yield spreads and interest rate movements: A bird’s eye view”. Review of Economic Studies 58, 495-514.

Carr, P., and L. Wu, (2003), ”The finite moment log stable process and option pricing”, Journal of Finance 58, 753-777

Carr, P., and L. Wu, (2004a), ”Time-changed Levy processes and option pricing”, Journal of Financial Economics 71,113-141.

Carr, P., and L. Wu, (2004b). ”Stochastic skew in currency options”. Working paper. New York University and Baruch College.

Carr, P. and L., Wu, (2009), ”Variance Risk Premiums” Review of Financial Studies, 22 (3): 1311-1341.

Carr, P., H. Geman, D. Madan, and M. Yor, (2002), ”The fine structure of asset returns: An empirical investigation”, Journal of Business 75, 305-332.

Carrasco, M., Chernov, M., Florens, J.-P., Ghysels, E., (2007). ”Efficient estimation of general dynamic models with a continuum of moment conditions”. Journal of Econometrics, in press, doi:10.1016/j.jeconom.

Carrasco, M., Florens, J.P., (2000): ”Generalization of GMM to a Continuum of Moment Conditions,” Econometric Theory, 16, 797-834.

Carrasco, M., Florens, J.P., (2008). ”On the asymptotic efficiency of GMM”. Manuscript, University of Rochester.

Chacko, G., and L.Viceira, (2003): ”Spectral GMM Estimation of Continuous-Time Processes,” Journal of Econometrics, 116, 259-292.

Chamberlain, G. (1987). ”Asymptotic efficiency in estimation with conditional moment restrictions”, Journal of Econometrics, 34, 305-334.

Chambers, J., C. Mallows, and B. Stuck, (1976), ”A method for simulating stable random variables”, Journal of the American Statistical Association 71, 340-344.

Chan, K.C., G.A. Karolyi, F.A. Longstaff, and A.B. Sanders, (1992), “An Empirical Comparison of Alternative Models of the Short-Term Interest Rate,” Journal of Finance, 47, 1209–1227.

Chapman, D., and N. Pearson, (2000), “Is the Short Rate Drift Actually Nonlinear?” Journal of Finance, 55, 355–388.

Chen, L., E. Bayraktar, and H. V. Poor (2005), ”Consistency Problems for Jump-Diffusion Models”, Appl. Math. Finance 12(2), 101–119.

Chen, S. X., Gao, J. and Tang, C. (2008). A test for model specification of diffusion processes. The Annals of Statistics 36, 167-198.

Chen, X., Hansen, L., and Carrasco, M. (2010). ”Nonlinearity and temporal dependence”, Journal of Econometrics, Volume 155, Issue 2, 155-169.

Chen, B. and Hong, Y. 2010, ”Characteristic function-based testing for multifactor continuous-time models via nonparametric regression”, forthcoming in Econometric Theory.

Chen, B. and Z., Song, (2010), ”Testing Whether the Underlying Continuous Time Process follows a Diffusion: an Infinitesimal Operator Based Approach,” manuscript.

Cheng, P., and Scaillet, O., (2007). ”Linear-quadratic jump-diffusion modeling”. Mathematical Finance 17, 575–598.

Cheridito, P., Filipovic, D., Kimmel, R.L., (2007), ”Market Price of Risk Specifications for Affine Models: Theory and Evidence”. Journal of Financial Economics. Vol. 83, no. 1. (January): 123-170.

Chernov, M., Gallant, A.R., Ghysels, E., Tauchen, G., (2003). ”Alternative models for stock price dynamics”. Journal of Econometrics 116, 225–257.

Cochrane, J. H., and M. Piazzesi. (2005), ”Bond Risk Premia”. American Economic Review 95:138–60.

Collin-Dufresne, P. and R.S. Goldstein, (2002), ”Do Bonds Span the Fixed Income Markets? Theory and Evidence for Unspanned Stochastic Volatility”, Journal of Finance 57, 1685-1729.

Collin-Dufresne, P. and R.S. Goldstein, and C. S. Jones. (2009), ”Can interest rate volatility be extracted from the cross section of bond yields?” Journal of Financial Economics, 94(1):47–66.

Conley, T.G., L.P. Hansen, E.G.J. Luttmer, and J.A. Scheinkman, (1997), “Short-Term Interest Rates as Subordinated Diffusions,” Review of Financial Studies, 10, 525–578.

Corradi, V. and N.R. Swanson. 2005. “Bootstrap specification tests for diffusion process”, Journal of Econometrics 124, 117-148.

Corradi, V. and H. White. 1999. “Specification Tests for the Variance of a Diffusion.” Journal of Time Series Analysis 20: 253-270.

Cox, J.C., J.E. Ingersoll, and S.A. Ross, (1981), “A reexamination of traditional hypotheses about the term structure of interest rates.” Journal of Finance 36, 321–346.

Cox, J.C., J.E. Ingersoll, and S.A. Ross, (1985), “A Theory of the Term Structure of Interest Rates,” Econometrica, 53, 385–407.

Czellar V., A. Karolyi, and E. Ronchetti, (2007), ”Indirect robust estimation of the short-term interest rate process”, Journal of Empirical Finance 14, 546–563

Dai, Q. and K. Singleton, (2000), ”Specification analysis of affine term structure models”. Journal of Finance 55, 1943—1978.

Dai, Q. and K. Singleton., (2002), “Expectation Puzzles, Time-Varying Risk Premia, and Dynamic Models of the Term Structure.” Journal of Financial Economics 63, 415–441.

Dai, Q., Singleton, K., (2003). ”Term structure dynamics in theory and reality”. Review of Financial Studies 16, 631–678.

Dai, Q., K. Singleton, and W. Yang, (2004). “Predictability of Bond Risk Premia and Affine Term Structure Models”, manuscript.

Das, S.R., (2002). ”The surprise element: Jumps in interest rates”. Journal of Econometrics 106, 27–65.

Davidson, J. (1994) Stochastic Limit Theory. Oxford: Oxford University Press.

Devroye, L. P., and T. J. Wagner, (1980): “Distribution-Free Consistency Results in Nonparametric Discrimination and Regression Function Estimation,” The Annals of Statistics, 8, 231–239.

Diebold, F.X., Hahn, J., Tay, A.S., (1999). ”Multivariate density forecast evaluation and calibration in financial risk management: high-frequency returns of foreign exchange”. Review of Economics and Statistics 81, 661–673.

Domínguez, M. A. and I. N. Lobato (2004). ”Consistent estimation of models defined by conditional moment restrictions”, Econometrica, 72, 1601-1615.

Donald, S. G., G. W. Imbens, and W. K. Newey (2003). ”Empirical likelihood estimation and consistent tests with conditional moment restrictions”, Journal of Econometrics, 117, 55-93.

Duarte, J. (2004). ”Evaluating an Alternative Risk Preference in Affine Term Structure Models”. Review of Financial Studies 17:379–404.

Duffee, G. R. (2002). ”Term premia and interest rate forecasts in affine models”. Journal of Finance 57, 405—443.

Duffie, D. and P. Glynn (2004). ”Estimation of continuous-time Markov processes sampled at random time intervals”. Econometrica 72 (6), 1773{1808.

Duffie, D., and R. Kan, (1996): ”A Yield-Factor Model of Interest Rates”, Math. Finance 6, 379–406.

Duffie, D., J. Pan, and K. Singleton (2000): ”Transform Analysis and Asset Pricing for Affine Jump-Diffusions,” Econometrica, 68, 1343-1376.

Duffie, D., L. Pedersen, and K. Singleton, 2003, “Modeling Credit Spreads on Sovereign Debt: A Case Study of Russian Bonds,” Journal of Finance, 55, 119–159.

Duffie, D, and K. Singleton, (1993), ”Simulated moments estimation of Markov models of asset prices”, Econometrica 61, 929–952.

Durham, G. B. and A. R. Gallant (2002). ”Numerical techniques for maximum likelihood estimation of continuous time diffusion processes”. Journal of Business and Economic statistics 20 (3), 297-338.

Durrett, R. (2005). “Probability: Theory and Examples”, 3rd Edition. Duxbury Press.

Dynkin, E.B. (1965), ”Markov Processes”, Volume I, Springer-Verlag. 32

Elerian, O., Chib, S. & Shephard, N. (2001). ”Likelihood inference for discretely observed non-linear diffusions”. Econometrica, 69, 959-993.

Egorov, A.V., Li, H., Xu, Y., (2003). ”Maximum likelihood estimation of time inhomogeneous diffusions”. Journal of Econometrics 114, 107–139.

Eraker, B. (2001). ”MCMC analysis of diffusion models with application to finance”. J. Bus. and Econom. Statist., 19, 177-191.

Fama, E. (1984a). ”The information in the term structure”. Journal of Financial Economics 13, 509-528.

Fama, E. (1984b). ”Term premiums in bond returns”. Journal of Financial Economics 13, 529-546.

Fama, E. (2006), ”The Behavior of Interest Rates.” The Review of Financial Studies, 2006, 19(2), pp. 359-79.

Fama, E. and R. Bliss (1987). ”The information in long-maturity forward rates”. American Economic Review 77 (4), 680-692.

Fan, Y. and Fan, J., 2008, ”Testing and detecting jumps based on a discretely observed process”, Manuscript.

Fan, Y., and Li, Q., (1999). ”Central limit theorem for degenerate U-statistics of absolutely regular processes with applications to model specification testing”. Journal of Nonparametric Statistics 10, 245.271.

Fan, J.and Wang, Y. (2007). Multi-scale jump and volatility analysis for high-Frequency financial data,Jour. Ameri. Statist. Assoc.,102,1349-1362.

Fan, J. & Zhang, C. 2003, ”A re-examination of Stanton’s diffusion estimation with applications to financial model validation”, Journal of the American Statistical Association 457, 118–134.

Fisher, I., (1896), “Appreciation and Interest,” American Economic Review Publications, vol. 11,, pp. 331– 442.

Gallant, A.R., and G. Tauchen, (1996), “Which Moments to Match?” Econometric Theory, 12, 657–681.

Gallant, A. R., and G. Tauchen, (1998), “Reprojecting Partially Observed Systems with Application to Interest Rate Diffusions,” Journal of American Statistical Association, 93, 10-24.

Gospodinov, N. and T., Otsu, (2009), ”Local GMM estimation of time series models with conditional moment restrictions”, forthcoming in Journal of Econometrics.

Gourieroux, C., Monfort, A. & Renault, E. (1993).” Indirect inference”. J. Appl. Econometrics, 8, S85-S118.

Hall, P., and C. Heyde, 1980, Martingale Limit Theory and Its Application, Academic Press, New York.

Hannan, E. 1970, Multiple Time Series, Wiley, New York.

Hansen, L.P. (1982). ”Large sample properties of generalized method of moments estimators”. Econometrica, 50, 1029-1054.

Hansen, B.E. (2008), ”Uniform convergence rates for kernel estimation with dependent data”. Econometric Theory 24, 726–748.

Hansen, L.P., and J.A. Scheinkman, (1995), “Back to the Future: Generating Moment Implications for Continuous Time Markov Processes,” Econometrica, 63, 767–804.

Hansen, L.P., and J.A. Scheinkman, 2003, ”Semigroup pricing”, manuscript.

Hansen, L.P., J.A. Scheinkman, and Nizar Touzi. (1998). ”Spectral methods for identifying scalar diffusions”. Journal of Econometrics, Volume 86, Issue 1, September, Pages 1-32

Heidari, M. and L. Wu, (2003), ”Are Interest Rate Derivatives Spanned by the Term Structure of Interest Rates?”, Journal of Fixed Income Market 13, 75-86

Heston, S., (1993). ”A closed-form solution for options with stochastic volatility with applications to bonds and currency options”. Review of Financial Studies 6, 327–343.

Hong, Y. (1999): ”Hypothesis Testing in Time Series via the Empirical Characteristic Function: a Generalized Spectral Density Approach,” Journal of the American Statistical Association, 94, 1201-1220

Hong, Y. and Lee, Y. 2005, “Generalized Spectral Tests for Conditional Mean Specification in Time Series with Conditional Heteroskedasticity of Unknown Form”, Review of Economic Studies, Vol. 72, 499-541

Hong, Y., Lee, Y. and Song, Z. 2008: ”Is the drift of the interest rate process linear? Evidence from a new approach”, working paper.

Hong, Y. and Li, H. (2005): ”Nonparametric Specification Testing for Continuous-Time Models with Applications to Term Structure of Interest Rates,” Review of Financial Studies, 18, 37-84.

Hong, Y., H. Li, and F. Zhao (2007) “Can the Random Walk be Beaten in Out-of-Sample Density Forecasts: Evidence from Intraday Foreign Exchange Rates,” Journal of Econometrics 141, 736-776.

Huang, J. and L.Wu, (2003), ”Specification analysis of option pricing models based on time-changed Levy processes”, Journal of Finance 59, 1405-1439.

Jarrow R. (1981). ”Liquidity Premiums and the Expectation Hypothesis”, J. Bank. Finance 5:539-546.

Jarrow R. (2009). ”The Term Structure of Interest Rates”, Annual Review of Financial Economics, 2009, 1:69-96.

Jarrow, R., Li, H., and Zhao, F., (2007). ”Interest rate caps ‘smile’ too! But can the LIBOR market models capture it? ” Journal of Finance 61, 341–378.

Jarrow R., P. Protter, and K. Shimbo, (2006), ”Asset price bubbles in a complete market”, Advances in Mathematical Finance, In Honor of Dilip B. Madan: 105 - 130.

Jarrow R., P. Protter, and K. Shimbo, (2010), ”Asset Price Bubbles in Incomplete Markets”, Mathematical Finance, 20 (2), (April 2010).

Jiang, G., and J. Knight, 1997, “A Nonparametric Approach to the Estimation of Diffusion Processes with an Application to a Short-Term Interest Rate Model,” Econometric Theory, 13, 615–645.

Jiang, G., and J. Knight, 2002, “Estimation of Continuous-Time Processes via the Empirical Characteristic Function,” Journal of Business and Economic Statistics, 20, 198–212.

Jiang G., and Yan S. (2009). ”Linear-quadratic term structure models-Toward the understanding of jumps in interest rates”, Journal of Banking and Finance, 33 (3), pp. 473-485.

Johannes, M., (2004). ”The statistical and economic role of jumps in continuous-time interest rate models”. Journal of Finance 59, 227–260.

Johannes, M. & Polson, N. (2009). ”MCMC methods for financial econometrics”. Handbook of Financial Econometrics. Amsterdam: North-Holland. Forthcoming.

Johannes, M. and Sundaresan, S. (2007), ”The Impact of Collateralization on Swap Rates”. The Journal of Finance, 62: 383–410.

Jorion, P., (1988), ”On Jump Processes in the Foreign Exchange and in the Stock Markets”, Review of Financial Studies 1, 427-455.

Joslin S. (2007), ”Pricing and Hedging Volatility Risk in Fixed Income Markets”, manuscript.

Kallenberg, O. (2002), ”Foundations of Modern Probability”, 2nd edition, Springer-Verlag.

Kanaya, S. (2007), ”Non-Parametric Specification Testing for Continuous-Time Markov Processes:Do the Processes Follow Diffusions?”, Manuscript, University of Wisconsin.

Kanaya, S and Kristensen, D. (2009), ”Estimation of Stochastic Volatility Models By Nonparametric Filtering”, working paper

Karatzas, I. und St. E. Shreve, (1991). ”Brownian motion and stochastic calculus”. Graduate Texts in Mathematics, 113 Springer-Verlag, New York.

Kent, J., (1978). ”Time reversible diffusions”. Advanced Applied Probability 10, 819–835.

Kessler, M. & Søensen, M. (1999). ”Estimating equations based on eigenfunctions for a discretely observed diffusion process”. Bernoulli, 5, 299-314.

Kitamura, Y. (1997). Empirical likelihood methods with weakly dependent processes, Annals of Statistics, 25, 2084-2102.

Kitamura, Y and M, Stutzer, (1997), ”An information-theoretic alternative to generalized method of moments estimation”, Econometrica 65, 861–874

Kitamura, Y., G. Tripathi, and H. Ahn (2004). "Empirical likelihood-based inference in conditional moment restriction models", Econometrica, 72, 1167-1714.

Kretschmer, U., Pigorsch, C., (2004). "EMM estimation of time-changed Levy processes". Working paper. University of Bonn and University of Munich.

Kristensen, D. 2004. "A Semi-parametric Single-Factor Model of the Term Structure" FMG Discussion Papers DP501, LSE.

Kristensen, D. 2008a, "Pseudo-Maximum-Likelihood Estimation in Two Classes of Semi-parametric Diffusion Models", working paper, Columbia University.

Kristensen, D. 2008b, "Nonparametric Estimation and Mis-specification Testing of Diffusion Models", working paper, Columbia University.

Ledoux, M & Talagrand, M. (1991) Probability in Banach Spaces. New York: Springer-Verlag.

Lee, S. and Mykland, P. 2008, "Jumps in financial markets: a new nonparametric test and jump dynamics", Review of financial studies

Leippold, M., and L. Wu, (2002), "Asset Pricing Under the Quadratic Class," Journal of Financial and Quantitative Analysis, 37, 271-295.

Leonov, V.P. and Shiryaev, A.N. 1959. "On a method of calculation of semi-invariants", (English translation). Theor. Prob. Appl. 4 319-329.

Li, F. (2007) "Testing the Parametric Specification of the Diffusion Function in a Diffusion Process," Econometric Theory, 23, 221-250.

Li, H. and Z. Song (2010), "Model Comparison Testing for Continuous-Time Finance Models", manuscript.

Li, H., M. Wells, and L. Yu, (2008), ”A Bayesian Analysis of Return Dynamics with Levy Jumps”, Review of Financial Studies 21, 2345-2378

Li, H., M.Wells, and L. Yu, (2009), ”Return Dynamics with Levy Jumps: Evidence from Stock and Option Prices”, Mathematical Finance forthcoming.

Li, H. and Zhao, F. (2006), ”Unspanned Stochastic Volatility: Evidence from Hedging Interest Rate Derivatives”. The Journal of Finance, 61: 341–378.

Litterman, R., J. Scheinkman, and L. Weiss, (1991), ”Volatility and the yield curve”, Journal of Fixed Income 1, 49-53.

Lo, A.W., 1988, “Maximum Likelihood Estimation of Generalized Ito Processes with Discretely Sampled Data,” Econometric Theory, 4, 231–247.

Lobato, I. 2002, ”A consistent test for the martingale difference hypothesis”, Econometric Reviews.

Longstaff, F. A. (2000). “Arbitrage and the expectations hypothesis.” Journal of Finance 55, 989–994.

Longstaff, F. A., and E. S. Schwartz, (1992), “Interest Rate Volatility and the Term Structure: A Two-Factor General Equilibrium Model,” Journal of Finance, 47, 1259-1282.

Lutz, Frederick A., (1940), ”The Structure of Interest Rates”, The Quarterly Journal of Economics 55, 36-63.

Madan, D., P. Carr, and E. Chang, (1998), ”The variance gamma process and option pricing”, European Finance Review 2, 79-105.

Merton, R., (1976), ”Option pricing when the underlying stock returns are discontinuous”, Journal of Financial Economics 3, 125-144.

Merton, R. C., (1980), ”On Estimating the Expected Return on the Market: An Exploratory Investigation”, Journal of Financial Economics, 8, 323–361.

Nelson D.B. (1990), “ARCH Models as Diffusion Approximations”, Journal of Econometrics, 45, 7–38.

Newey, W. K. (1990). ”Efficient instrumental variables estimation of nonlinear models”, Econometrica, 58, 809-837.

Newey, W. K. (1993). ”Efficient estimation of models with conditional moment restrictions”, in G. Maddala, C. Rao, and H. Vinod (eds.), Handbook of Statistics, 11, pp. 419-454. Elsevier Science.

Newey, W.K., and R, Smith, (2004),”Higher order properties of GMM and generalized empirical likelihood estimators”, Econometrica 72, 219–256.

Newey, W. K. and D. McFadden (1994). ”Large sample estimation and hypothesis test”, Handbook of Econometrics, Vol. IV, pp. 2111-2245. Elsevier Science.

Øksendal B. (2003). ”Stochastic Differential Equations: An Introduction with Applications”. Springer-Verlag (Universitext), Berlin-Heidelberg-New York (6th ed.).

Pan, J., (2002). ”The jump-risk premia implicit in options: evidence from an integrated time-series study”. Journal of Financial Economics 63, 3–50.

Park, J. 2008, ”Martingale regression and time change”, working paper.

Park, J. and Whang, Y.J. 2003, ”Testing for the martingale hypothesis”, working paper, Department of Economics, Rice University and Department of Economics, Korea University.

Parthasarathy, K.P., (1967). ”Probability Measures on Metric Spaces”. New York-London.

Pedersen, A. R. (1995), “A New Approach to Maximum Likelihood Estimation for Stochastic Differential Equations Based on Discrete Observations,” Scandinavian Journal of Statistics, 22, 55–71.

Phillips, P.C.B., (1973), ”The Problem of Identi cation in Finite Parameter Continuous Time Models”, Journal of Econometrics, 4, 351-262.

Phillips, P.C.B. and J. Yu, (2005), ”Jackknifing Bond Option Prices”. Review of Financial Studies, 18, 707-742.

Phillips, P.C.B. and Yu, J. (2009). ”A Two-Stage Realized Volatility Approach to Estimation of Diffusion Processes with Discrete Data”, Journal of Econometrics, 150, 139-150.

Piazzesi, M., (2003), ”An Econometric Model of the Yield Curve with Macroeconomic Jump Effects”, NBER Working Paper 8246.

Piazzesi, M., (2005). ”Bond yields and the Federal Reserve”, Journal of Political Economy Volume 113, Issue 2, pp. 311-344.

Piazzesi, M., (2010). ”Affine term structure models”. Forthcoming in the Handbook of Financial Econometrics.

Priestley, M.B. (1981). ”Spectral Analysis and Time Series”. London: Academic Press.

Pritsker, M., (1998), “Nonparametric Density Estimation and Tests of Continuous Time Interest Rate Models,” Review of Financial Studies, 11, 449–487.

Protter, P. (2005), ”Stochastic Integration and Differential Equations, Second Edition”, Version 2. Springer-Verlag, Heidelberg.

Qin, J. and J. Lawless (1994). Empirical likelihood and generalized estimating equations, Annals of Statistics, 22, 300-325.

Ranga Rao, R. (1962): ”Relations Between Weak and Uniform Convergence of Measures with Applications,” Annals of Mathematical Statistics, 33, 659-680.

Revuz, D and Yor, M. (2005), ”Continuous martingales and Brownian motion”, volume 293 of Grundlehren der Mathematischen Wissenschaften. Springer-Verlag, Berlin.

Roberds,W. and C. Whiteman (1999). ”Endogenous term premia and anamolies in the term structure of interest rates: Explaining the predictability smile”. Journal of Monetary Economics 44, 555-580.

Robinson, P. (1987): “Asymptotically Efficient Estimation in the Presence of Heteroscedasticity of Unknown Form” Econometrica 55 875-891.

Robinson, P. (1991): “Best Nonlinear Three-Stage Least Squares Estimation of Certain Econometric Models,” Econometrica, 59, 755–786.

Rogers, L.C.G., and Williams, D. (2000), ”Diffusions, Markov Processes and Martingales”, Vol. 1, 2nd edition, Cambridge University Press.

Sam, G. and Jiang, G. 2007. “Nonparametric Estimation of the Short Rate Diffusion from a Panel of Yields,” forthcoming in Journal of Financial and Quantitative Analysis.

Santa-Clara, P., (1995). ”Simulated likelihood estimation of diffusions with an application to the short term interest rate”. Ph.D. Dissertation, INSEAD.

Schaumburg, E., (2001). ”Maximum likelihood estimation of lévy type SDEs”. Ph.D. Thesis, Princeton University.

Schoutens, W. 2003. ”Levy Processes in Finance”. Wiley, New York.

Serfling, R. J. (1980): Approximation Theorems of Mathematical Statistics. John Wiley.

Shreve, S., (2004), ”Stochastic Calculus for Finance II: Continuous-Time Models”, Springer Verlag, ISBN 0-387-40101-8.

Singleton, K., 2001, “Estimation of Affine Asset Pricing Models Using the Empirical Characteristic Function,” Journal of Econometrics, 102, 111–141.

Singleton, K., (2006), ”Empirical Dynamic Asset Pricing”, Princeton University Press.

Smith, R.J., (2007). ”Efficient information theoretic inference for conditional moment restrictions”. Journal of Econometrics 138, 430-460.

Song, Z. (2011). ”A Martingale Approach for Testing Diffusion Models Based on Infinitesimal Operator” Journal of Econometrics, 162-2 , 189-212.

Stambaugh, Robert F., (1988). ”The information in forward rates : Implications for models of the term structure,” Journal of Financial Economics, vol. 21(1), pages 41-70

Stanton, R., (1997), “A Nonparametric Model of Term Structure Dynamics and the Market Price of Interest Rate Risk,” Journal of Finance, 52, 1973–2002.

Stein, E.M., Stein, J.C., (1991). Stock price distributions with stochastic volatility: an analytic approach. Review of Financial Studies 4, 727–752.

Stratonovich, R.L., 1963. ”Topics in the Theory of Random Noise”, 1, (English translation). Gordon and Breach, New York.

Stroock, D.W. and Varadhan, S.R.S. (1969), ”Diffusion processes with continuous coefficients I & II”, Comm. Pure & Appl. Math. 22, 345-400 and 479-530.

Stute, W. (1997): “Nonparametric Model Checks for Regression,” The Annals of Statistics, 25, 613–641.

Sundaresan, S., (2000), "Continuous-Time Methods in Finance: A Review and an Assessment," Journal of Finance, 55, 1569–1622.

Takamizawa, H. 2008: "Is Nonlinear Drift Implied by the Short End of the Term Structure?" Review of Financial Studies, vol. 21, issue 1, pages 311-346

Tang, C.Y. and S.X. Chen, (2009), "Parameter Estimation and Bias Correction for Diffusion Processes". Journal of Econometrics, forthcoming.

Tanaka, K. (1996): Time Series Analysis. New York: Wiley and Sons.

Tauchen, G., 1997, "New Minimum Chi-Square Methods in Empirical Finance," in D. Kreps and K.Wallis (eds.), Advances in Econometrics: Seventh World Congress, Cambridge University Press, Cambridge, UK.

Thompson S. (2008), "Identifying term structure volatility from the LIBOR-swap curve Review of Financial Studies", 21 (2), pp. 819-854.

Tripathi, G., and Kitamura, Y., (2003). "Testing conditional moment restrictions". Annals of Statistics 31, 2059–2095.

Vasicek, O., (1977), "An Equilibrium Characterization of the Term Structure," Journal of Financial Economics, 5,

Wong, E. (1964). "The construction of a class of stationary Markov processes". In R. Bellman (Ed.), Sixteenth Symposium in Applied Mathematics - Stochastic Processes in Mathematical Physics and Engineering,pp. 264-276. Providence, RI: American Mathematical Society.

Wu, L. (2008), "Modeling Financial Security Returns Using Levy Processes", Handbooks in Operations Research and Management Science: Financial Engineering, 15, Eds. John Birge and Vadim Linetsky, Elsevier.

Yu, J. (2007), “Closed-Form Likelihood Approximation and Estimation of Jump-Diffusions With an Application to the Realignment Risk Premium of the Chinese Yuan,”Journal of Econometrics, 141, 1245–1280.

Zähle, H. 2008, ”Weak approximations of SDEs by discrete-time processes”, Journal of Applied Mathematics and Stochastic Analysis.

Zhang, L. (2008). ”Estimating Covariation: Epps Effect, Microstructure Noise”, working paper, University of Illinois at Chicago.

Zhou H, (2001). ”Jump-diffusion term structure and Ito conditional moment generator,” working paper.